CORRESPONDENCE BETWEEN AQUATIC ECOREGIONS AND THE DISTRIBUTION OF FISH COMMUNITIES OF EASTERN OKLAHOMA

Charles E. Howell, Jr. B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

May 2001

APPROVED:

William T. Waller, Major Professor
James H. Kennedy, Committee Member
Thomas W. LaPoint, Committee Member
Kenneth L. Dickson, Committee Member
Earl Zimmerman, Chair of the Department of Biological Sciences
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

Howell, Charles E., Jr. Correspondence between aquatic ecoregions and the distribution of fish communities of eastern Oklahoma. Master of Science (Environmental Science), May 2001, 57 pp., 2 tables, 13 figures, references, 82 titles.

I assessed fish community data collected by the Oklahoma Conservation Commission from 82 minimally impaired wadeable reference streams in eastern Oklahoma to determine whether existing aquatic ecoregions provide the best framework for spatial classification for the development of biological assessment methods and biocriteria. I used indirect ordination and classification to identify groups of sites that support similar fish communities. Although correspondence was observed between fish assemblages and three montane ecoregions, the classification system must be refined and expanded to include major drainage basins and physical habitat attributes for some areas to adequately partition variance in key measures of biological integrity. Results from canonical correspondence analysis indicated that substrate size and habitat type were the primary physical habitat variables that influenced the fish species composition and community structure.

ACKNOWLEDGMENTS

I thank my committee members, William T. Waller, James H. Kennedy, Thomas W. LaPoint, and Kenneth L. Dickson for their help with my thesis, and for their many contributions to the development and refinement of biological monitoring and assessment methods. John Hassell and Scott Stoodley, former and current Directors of the Oklahoma Conservation Commission Water Quality Program, deserve recognition for supporting their staff's efforts to collect data needed to characterize reference conditions. Dan Butler, Oklahoma Conservation Commission, selected the sites, oversaw the data collection, assisted with reference site selection, and provided much insight from the perspective of an experienced field biologist. Philip Moershel, Kendra Eddlemon, and Derek Johnson patiently answered my questions about their data and database, and provided the agency's standard operating procedures. Schad Meldrum developed the relational database. The project would not have been successful without the hard work of dedicated field sampling crews with the Commission, including Wes Shockley, Brooks Trammell, and many others. Chuck Potts, Oklahoma Water Resources Board, provided input on the project preproposal. Tom Nelson, U.S. Environmental Protection Agency Region 6, assisted in locating the geographic information system data layers. Valério De Patta Pillar, Universidade Federal do Rio Grande do Sul, Brazil, provided advice for implementing his approach to identify significant clusters. Mike Palmer, Oklahoma State University, deserves recognition for developing the ordination web pages, an effective mechanism for overcoming the barriers imposed by the terminology used in numerical ecology.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF ILLUSTRATIONS

INTRODUCTION

In 1987, the Oklahoma Conservation Commission (hereinafter "Commission") faced the responsibility of producing an assessment of the nature and extent of nonpoint source pollution throughout the state. The basis for the initial assessment was largely limited to water chemistry data collected under low flow conditions from an historic fixed station network, and visual observations made by field staff. In 1991, the Commission initiated a comprehensive monitoring program to characterize the relationships between water quality, biological integrity, and physical habitat conditions of wadeable streams, in an effort to improve the nonpoint source assessment.

The benefits of including biological data in the monitoring program are well documented. Biological monitoring is an explicit requirement of section 106 of the Clean Water Act. The approach provides a direct measure of progress toward a primary goal of the Act – to restore and maintain the biological integrity of the Nation's waters (Mount 1994, Karr and Chu 1999, Yoder and Rankin 1995). The results from biological monitoring represent the summation of all stressors acting upon aquatic communities (Anderson et al. 1995, Frenzel and Swanson 1996, Jester et al. 1992, Karr 1986, Petersen 1998, Wang et al. 1997). Biological monitoring often reveals problems that would remain undetected if data collection was limited to water column chemistry (Maxted 1997). Aquatic communities also integrate the effects of water quality conditions over time, providing a less variable and more cost-effective indicator than water column chemistry (Ohio EPA 1987, Karr and Chu 1999). Therefore, biological monitoring may be

conducted under stable, low flow conditions, avoiding the complications of collecting data in response to rainfall events.

Karr and Dudley (1981) proposed the current, widely accepted definition of biological integrity, as the ability of a waterbody to support and maintain "a balanced, integrated, adaptive community of organisms having a species composition, diversity, and functional organization comparable to that of natural habitat of the region." Reflection on the definition reveals the initial steps required for the development of biological assessment methods and biocriteria. First, a classification system must be established to identify the "natural habitat of the region" and minimize the spatiotemporal variability in measurable aquatic community attributes that represent the composition, diversity, and function of communities inhabiting undisturbed or least disturbed reference streams. Second, biological monitoring must be conducted over time at minimally impaired reference streams within these regions to quantify the range of variability observed in these measures (Barbour et al. 1999, Conquest 1993, Hirst 1984, Hughes et al. 1993, Karr 1993, Karr and Chu 1997, Omernik and Griffith 1991, Polls 1994, Voshell et al. 1997). Some recognized these steps as an iterative process that may be continually refined over time, as additional data and information become available (Hughes et al. 1994 and Grumbine 1994).

Ecological regions or ecoregions have become widely accepted as a starting point for the classification of streams for biocriteria development. Landscape scale factors largely determine the physical habitat conditions and water quality in streams (Frissell et al. 1986, Rohm et al. 1987). Bailey (1982) recognized ecological regions as essential to any resource management effort, and as an essential component in the design of cost-

effective sampling programs. Ecoregions provide a spatial basis for the compilation of data from many similar sites into a reference data set (Hughes and Larsen 1988). Streams within ecoregions generally respond in like manner to similar management practices or similar environmental stresses (Bailey 1982, Lyons 1989), although within region heterogeneity in physical habitat and water quality conditions may confound measurement of these responses (Toepfer et al. 1998).

As described by Karr and Chu (1999), the challenge in stream classification is to create a system with only as many classes as are needed to detect and describe the biological effects of human activity. If the classes are too broad, encompassing a greater range of natural variability, the biological assessment methods may lack the sensitivity needed to provide an adequate level of protection. If the classes are too narrow, the costs for characterizing reference conditions increase, because of the need to characterize additional classes.

Hawkes et al. (1986) suggested that contiguous fish ecoregions are useful for management, whereas areas with interspersed sites are inconvenient; although, interspersed sites or regions may be required where local geology is highly variable. Spindler (1986) identified a need for discontinuous ecoregions in Arizona where elevation appeared to be a critical factor in determining macroinvertebrate community composition.

Hughes et al. (1993) suggested that ecoregions should be evaluated, based on the response of multiple assemblages to avoid the development of assemblage-specific maps. They described that such maps would be difficult for state and federal land managers to use. In contrast, Commission staff expressed interest in evaluating fish and

macroinvertebrate assemblages independently to ensure that appropriate reference streams are identified for each taxonomic group.

Description of Existing Aquatic Ecoregions

Jarman (1984) and Omernik (1987) concurrently developed maps of aquatic ecoregions of Oklahoma. Omernik (1987) relied on patterns in land surface form, potential natural vegetation, soil types, and land use, and identified 7 ecoregions in the eastern part of Oklahoma (Figure 1). Jarman (1984) considered these factors, and added rainfall and runoff, watershed area, evapotranspiration, and watershed slope (Jarman 1984). Their results were similar for most of the state, although Jarman (1984) grouped portions of Omernik's ecoregions – the Ozark Highlands, Boston Mountains, the Arkansas Valley, and Ouachita Mountains – into 2 regions. The cartographers described regions that were heavily influenced by the east to west variation in climate, ranging from the humid east to the semiarid west, as well as terrain that ranges from mountainous areas of the east to sandy flats in the west. Local variations in soil types and parent rock material contribute to the definition of the ecoregions and variability within the ecoregions, because of their effects on water quality, physical habitat conditions, and potential natural vegetation.

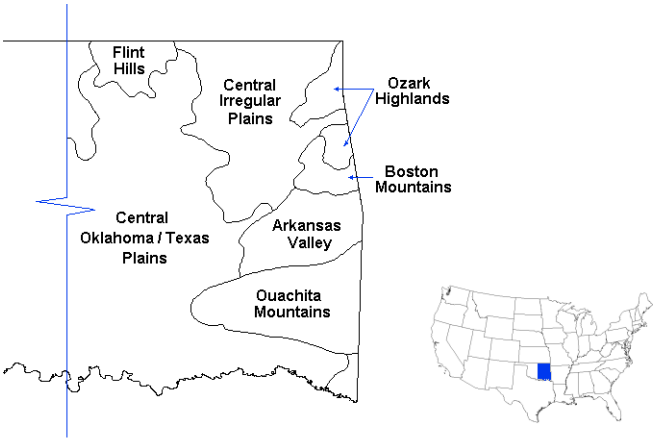


Figure 1. Study area covering the eastern half of Oklahoma, and ecoregions described by Omernik (1987).

Why Evaluate Ecoregions?

The U.S. Environmental Protection Agency's (EPA) Science Advisory Board has described the need to refine ecoregion classification techniques, because the distribution of aquatic organisms may not coincide with existing ecoregion boundaries (EPA 1993). Similarly, many researchers have described the need to evaluate the effectiveness of any classification scheme, before using it for management and reporting of results (Bailey 1982, Gallant et al. 1989, Hughes and Larsen 1988, Karr and Chu 1997, Omernik and Griffith 1991). Some researchers have observed considerable correspondence between existing aquatic ecoregions and fish assemblages, although most acknowledge that additional factors must be taken into account to adequately classify streams (Hughes et al. 1993, Lyons 1989, Rohm et al. 1987, Toepfer et al. 1998). Others have identified the need to delineate subregions and combine portions of ecoregions to better reflect natural variability of aquatic assemblages (Dewalt 1995, Hawkes et al. 1986, Hornig et al. 1987, Spindler 1996).

I assessed fish community data from reference quality streams collected by the Commission from eastern Oklahoma to determine whether existing aquatic ecoregions (Jarman 1984, Omernik 1987) provide an adequate framework for spatial classification of wadeable streams and rivers for the purpose of biocriteria development. The primary goal of the study was to identify aquatic ecoregions that minimize spatial variability in measures of fish community integrity for reference quality streams. I also used canonical correspondence analysis to identify the primary physical habitat variables that influence the species composition and structure of reference stream fish assemblages of eastern Oklahoma, as an initial step toward the development of models to predict species

composition. This will allow for the development of bioassessment methods and criteria that are sensitive to subtle impairment and accurate in terms of limited response to natural variation.

METHODS

DATA COLLECTION

Site Selection

The Commission selected stream sites subjectively to include all potential reference quality streams, as well as others that may be impaired by poor water quality, degraded physical habitat conditions, or both, to ensure that an adequate data set would be available for the eventual development and evaluation of bioassessment methods (Karr and Chu 1999). For the purpose of this study, I selected a subset of the sites that were limited to include only minimally impaired reference streams. The intent was to select sites with a relatively intact riparian zone, stable banks, cattle exclusion, or limited access, and no known discharges or other potential sources of impairment.

Site-specific physical habitat measurements made by Commission staff, described below, provided the primary basis for reference site selections, although these data were available for only about half of the sites. I included streams with riparian zones that were characterized as predominantly either “stable forest” or “good condition grasslands” and “moderately used forest” or “fair condition grassland”. Notes made by Commission staff about potential sources of impairment were a key component of this review.

ESRI ArcView version 3.2 was used to screen candidate reference sites for proximity to potential sources of impairment and areas of heavy land use. Point data layers included locations of National Pollutant Discharge Elimination System dischargers from the EPA Permit Compliance System, sites registered as Resource Conservation and Recovery Act facilities, "Superfund" facilities from the National Priorities List, the EPA Toxics Release Inventory, and locations of other major federal facilities. I also evaluated the proximity of sites to areas where ambient toxicity has been observed though a screening program conducted by Oklahoma water quality agencies and the EPA (1997).

Land use and land cover pattern maps that depict conditions between 1991 and 1993, at a resolution of 30 meters per pixel (Riitters et al. 2000), were a key component of the geographic information system (GIS) data review. The maps were used to identify areas of intensive land use, such as row crop agriculture and urban uses, and the potential for reduction or elimination of riparian zones where land uses eliminate native forests or grasslands.

After the initial screening of physical habitat data and GIS data layers, I excluded sites where fish collection efforts yielded less than 150 individuals or fewer than 8 species. This step eliminated some true reference quality streams that are naturally faunal poor (Dan Butler, personal communication), but I believed these collections would contribute limited information to the study and be a potential source of inordinate variability, based on the findings of Fore et al. (1994).

I also classified sites by existing aquatic ecoregions and watershed size to allow comparison of key metrics of fish community integrity, primarily species richness. I calculated descriptive statistics for the distribution of the numbers of individuals and

species collected within each class of reference streams to allow the observation and exclusion of outliers, defined as any value falling below the 25th percentile minus 1.5 times the interquartile range. Last, I relied heavily on the knowledge and advice of experienced Commission field biologists, Dan Butler and Derek Johnson, in final decisions about streams that represent minimally impaired conditions. The reference streams and fish collections that were included in this study were listed in Appendix B.

Fish Collection

The Commission described their data collection procedures in detail in their standard operating procedures (OCC 1996). Fish collections were made along a representative 400-meter stream reach, under stable low flow conditions. Most collections were made during the months of June through September with some collections made in October and November.

Field sampling crews typically used both seining and electrofishing and recorded the collections separately for each method, unless site-specific conditions precluded the use of either method. During the first year of the program, in 1991, the Commission employed a limited level of sampling effort by electrofishing for a total of 15 minutes, as indicated by a timer on the backpack electrofisher. Concerns about the representativeness of this level of sampling effort led them to adopt a more rigorous approach in subsequent years. Seining was conducted first in a downstream direction using seines of varying lengths, dependent on the size of the stream being sampled. A variety of seining techniques was used, dependent on the habitat types encountered. Sampling of all available habitat types continued until no new species were collected on subsequent hauls. Electrofishing was then used primarily to sample habitat types that could not be

seined effectively, like brush piles, roots, and cobble substrates. Again, sampling continued throughout the reach until no new species were collected on subsequent attempts. Once fish collection was completed, larger fish were identified in the field and returned to the stream. Smaller fish were placed in 10% formalin and returned to the laboratory for identification and enumeration.

Physical Habitat Characterization

Physical habitat observations and measurements were made at 20 equally spaced transects along each 400 meter reach (OCC 1996). These included channel morphology measurements, substrate size class and embeddedness estimates, observations of deposition and scouring, observations of instream habitat type and instream cover, canopy cover estimates, notes on bank stability and bank vegetative protection, and riparian buffer zone width and condition. Commission staff also made several unique observations to identify and, in some cases, quantify the potential for impairment resulting from stream bank destabilization or other activities that may result in an influx of sediments or nutrients. These included the presence of cattle exclusion or fencing, evidence of livestock trampling, the presence and number of manure piles, the number of cattle trails crossing the stream and size class of each trail, the presence of cattle within the riparian zone, the presence of gravel roads that may contribute sediments, and the presence of discharges from pipes.

DATA ANALYSIS

After identifying candidate reference sites, I used complementary ordination and classification techniques to elucidate groups of sites supporting similar fish communities. Kenkel and Orlóci (1986) evaluated several different ordination approaches and found

nonmetric multidimensional scaling (NMDS) to be the best approach for recovering simulated coenoplanes. They recognized detrended correspondence analysis (DCA) as the most successful metric approach and recommended it as a complement to NMDS. DCA has been widely used for our purpose and interest in NMDS appears to be growing in the published literature (Jongmann 1995, Legendre and Legendre 1993, Matthews et al. 1992, Palmer 2000, Tetra Tech 2000). I also used cluster analyses, primarily as an aid in interpreting the ordinations. Last, I used canonical correspondence analysis (CCA) to examine the relationships between selected environmental variables and reference stream fish assemblages.

Selection of a Distance Measure and Clustering Algorithm

The selection of a measure of ecological distance or dissimilarity between fish communities at paired sites is a critical aspect of applying multidimensional scaling and classification. I avoided distance metrics that reduce species counts to binary presence/absence (Gallant et al. 1989, Hawkes et al. 1986). Although the reduction to presence/absence greatly reduces variability, it also eliminates the ability to assess differences in the relative abundance of each species, and the potential for a finer degree of resolution between sites (Echelle and Schnell 1976). The composition and relative abundance of species provide the basis for many biological assessment metrics; therefore, it is imperative to integrate this information into the effort to refine the spatial classification.

Ecological distance measures differ in their ability to distinguish minor differences in community composition, and some may be inordinately influenced by sampling variability (Boyle et al. 1990, Cao et al. 1997a, Legendre and Legendre 1993).

Some of the most popular distance measures do not fare well when tested with simulated data and some that work well with simulated data have not been used extensively in field studies (Boyle et al. 1990, Cao et al. 1997a). Cao et al. (1997a) recently formulated a robust measure of dissimilarity, dubbed CYd, that places the greatest weight on differences in species counts that may be attributable either to loss of a species or shifts in the abundance of a species, while placing lesser weight on differences that may be insignificant or attributable to sampling variability. Data transformation is not required, because the metric does not possess the inherent bias observed in other commonly used metrics. Cao et al. (1997a) demonstrated success in grouping replicate samples collected from the same sites while discriminating between sites with only minor differences in water quality using the CYd metric with classification and multidimensional scaling. The metric may be calculated, as follows:

$$CYd = \frac{1}{n}\sum \frac{(Xij + Xkj)\log 10\left(\frac{Xij + Xkj}{2}\right) - Xij \log 10 Xkj - Xkj \log 10 Xij}{Xij + Xkj}$$

where n is the total number of species present in both samples, Xij is the number of individuals of species j in sample i, and Xkj is the number of individuals of species j in sample k. I modified the Basic code developed by Ludwig and Reynolds (1988) to calculate the CYd distance between every possible combination of paired sites within Microsoft Excel (Appendix A).

The objective to identify broad groups of reference sites supporting similar fish communities led to the selection of Ward's method or minimum variance algorithm for clustering. Rather than focusing on distances between clusters, the method determines how much variation is within each cluster and adds samples that least increase this

variation (Legendre and Legendre 1993). Ward's method and complete linkage were the most successful clustering algorithms evaluated by Cao et al. (1997b). For reasons described in the next section, square root transformed counts were also clustered by Ward's method by squared Euclidean distances.

Number of Significant Clusters

Cluster analyses always result in the formation of clusters or groups, even when meaningful or significant partitions do not exist. The same general problem also applies when interpreting ordination biplots, and subjective decisions are sometimes required about which samples on a biplot represent a group. This appeared to present an obstacle in applying these approaches to identify unique subregions, based primarily on samples of aquatic communities. Milligan and Cooper (1985) compared 30 different approaches to determine the number of clusters in a dataset, but caveat their recommendations by stating that the apparent success of some approaches was probably dependent on the structure of the data used in the comparisons. In addition, they did not attempt to identify the power of each approach, and the simulated datasets used in their comparison had well-defined clusters (Pillar 1999). The problem of significant cluster identification is the subject of much ongoing research.

In an effort to develop widely applicable methods, researchers have applied bootstrap resampling to evaluate the significance of clusters (Legendre and Legendre 1998, Nemec and Brinkhurst 1988, Pillar 1999). Efron (1981) and Scheiner (1993) suggested that resampling techniques, such as the bootstrap, may be the most appropriate choice for data analysis when the distribution of a test statistic is unusual or unknown. Nemec and Brinkhurst (1988) described a bootstrap method to evaluate the significance

of clusters in a dataset, but their approach requires multiple replicate samples from each site. I did not have true replicate samples in the Oklahoma dataset, although some sites have been sampled more than once in different years. This is a typical problem in assessments of stream fish assemblages.

Pillar (1999) developed a bootstrap approach that does not require replicate samples to identify significant clusters. First, the samples are clustered by squared Euclidean distance and the user's choice of clustering algorithm. Then, the original data set is resampled with replacement, and the classification is recalculated many times, followed by a comparison between the original classification and each bootstrap classification. The approach is based on the assumption that "sharp" partitions will be present in most bootstrap classifications, whereas "fuzzy" partitions will not. Pillar (1999) validated the approach using simulated data and described an application using actual data. The test is carried out by specifying a number of clusters to evaluate, based on the following hypotheses:

Ho: The objects in the bootstrap clusters are random samples of objects in the corresponding clusters formed in the original classification, i.e., the specified numbers of partitions are "sharp".

Ha: The objects in the bootstrap clusters are not random samples of objects in the corresponding clusters formed in the original classification, i.e., the specified numbers of partitions are "fuzzy".

The method computes a test statistic and probability that the null hypothesis is true. I used Pillar's (1999) approach and software to identify the number of statistically significant clusters of sites in the fish assemblage data, selecting options for Ward's

minimum variance clustering algorithm, 1000 bootstrap classifications, and an alpha of 0.10. I attempted to modify Pillar's source code to use the CYd distance measure, rather than squared Euclidean distance. However, the modified application yielded probability values that were too low and failed to identify clusters that were known to exist in test data sets. I used an evaluation version of the Multivariate Statistical Program (MVSP) version 3.12b (Kovach Computing Services 2000) to verify that a CYd matrix sometimes yields negative eigenvalues when analyzed by principal coordinates analysis, an indication that the metric does not meet the axiom of triangle inequality and, therefore, is not a Euclidean metric (Legendre and Legendre 1993). Pillar (personal communication) confirmed my findings by testing my source code in his multivariate statistical program (MULTIV). Therefore, I used Pillar's software without modification.

Detrended Correspondence Analysis

A preliminary correspondence analysis of the fish abundance data, conducted using CANOCO for Windows version 4.0 (ter Braak and Šmilauer 1998), yielded a total inertia greater than 5.0, and a strong arch effect was observed in the ordination biplot, indicating that most of the fish species exhibited unimodal distributions along the ordination axes (ter Braak and Šmilauer 1998). Therefore, detrending by segments, a unimodal variant of correspondence analysis was used to group samples or sites supporting similar fish assemblages, based on the chi-square distance preserved in the analysis. DCA positions both species and sites simultaneously through an iterative approach referred to as "reciprocal averaging". Therefore, ordination biplots will reveal not only groups of sites, but also the species that influence the arrangement of sites.

There was no consensus in the literature, regarding the benefits of data transformations applied to species counts, before conducting DCA, probably because the most useful transformation is dependent on the structure of a specific dataset and the study objectives. Transformations were often applied to minimize variance and improve the interpretability of ordinations, and to reduce the effects of either abundant or rare species (Cao et al. 1987, Frenzel and Swanson 1996, Gallant et al. 1989, Hornig et al. 1994, Hughes 1984, Lyons 1989, Richards and Host 1993, Spindler 1996). I avoided a commonly used transformation that involves conversion of counts to proportional or percentage-type data, because Jackson (1997) demonstrated that it is possible to introduce artificial relationships into a data matrix that are predictable artifacts of the transformation. Such a transformation is unnecessary for use with DCA, because the chi-square metric inherently relies on relative abundance. Jackson (1997), Palmer (2000), and Lepš and Šmilauer (1999) suggested that correspondence analyses perform reasonably well without transforming the original abundance data. Through trial and error application of commonly used transformations and raw counts in the DCA, I found that square root transformation of fish species counts produced an interpretable ordination. Therefore, I applied square root transformations in all analyses that required calculation of either chi-square distance or Euclidean distance.

Gauch (1982) recommended the removal of rare taxa that are present at less than 5 percent of the sites before conducting DCA, because they may have an excessive influence on the ordination. The approach was often followed by ecologists (Hornig et al. 1994, Lyons 1989, Somers et al. 1998). However, Cao et al. (1998) and Karr and Chu (1999) made convincing arguments that excluding rare species will reduce the sensitivity

of community-based assessment methods. It is plausible that rare species observed at 5 percent of available reference sites may represent a unique subregion. The removal of rare species and down weighting of rare species dispersed sites in the DCA ordination, rather than improve its interpretability; therefore, I did not remove species in my final analyses.

Two samples were removed from all ordination and cluster analyses (nos. 12076 and 12077) because they were extreme outliers that distorted the ordinations. Both samples were collected from an unnamed tributary of Red Oak Creek (Appendix C).

Nonmetric Multidimensional Scaling

NMDS yields an ordination of a single selected measure of ecological distance or dissimilarity applied to every possible combination or pairing of sites (Kachigan 1991). Although it is impossible to accurately represent the ecological distances by the depicted distances in an ordination, after reduction to fewer dimensions, the amount of distortion may be expressed as a loss function or stress function (Jongmann et al. 1995, StatSoft, Inc. 1996). Stress values range upward from zero, which indicates a perfect match, and values less than 0.15 are generally considered acceptable (Kachigan 1991). The researcher must specify the number of ordination axes and supply an initial ordination as a starting point. I used Statistica for Windows version 5.1, which automates the initial ordination by conducting a principal components analysis. Additional NMDS ordinations were then conducted using the initial solution as the starting point to recheck the final solutions (StatSoft, Inc. 1996). I created a scree plot to observe the effects of dimensional reduction, and tried many different ordinations after specifying different numbers of dimensions, in attempts to find an interpretable solution (StatSoft, Inc. 1996).

It was necessary to eliminate 7 samples from the NMDS ordination, because Statistica limits the number of samples to 90 for this analysis. I deselected 5 samples from sites with multiple collections, in addition to removing 2 samples that distorted the DCA ordination (nos. 12076 and 12077).

Resolution of Differences between the Indirect Ordinations

NMDS, coupled with the CYd metric, appeared to offer the most straightforward approach to ordination or site grouping with the least potential to yield misleading results. DCA has been criticized for potential limitations, associated primarily with the required detrending and rescaling (Palmer 1988, Wartenburger et al. 1987), and limitations associated with the chi-square distance measure (Legendre and Legendre 1998). The outcome of cluster analyses may be affected by the selection of distance measures, clustering algorithms, and differences in samples included in the classification. Therefore, the results of NMDS were the primary basis for decisions to assign individual sites to an ecoregion. After conducting the individual analyses, I transferred groups of sites supporting similar assemblages and unique stations from the ordination biplots to maps for visual comparison with aquatic ecoregions. This approach integrated the wealth of existing information from physical classifications with the results from actual fish assemblage data (EPA 1991, Tetra Tech 2000).

Calculation of Key Bioassessment Metrics

I classified sites by the refined or redefined ecoregions, and calculated descriptive statistics for key biological assessment metrics to observe within and between region variability. The metrics included fish species richness, CYd distance (Cao et al. 1997a),

and weighted average tolerance indices, using the approach of Hilsenhoff (1982) with water quality and habitat quality tolerance values published by Jester et al. (1992).

Canonical Correspondence Analysis

The effectiveness of ecoregions, as a classification layer, may be limited by within-region heterogeneity in physical habitat and water quality conditions (Toepfer et al. 1998). For example, Echelle and Schnell (1976) described 4 unique assemblages within the Kiamichi River basin of Oklahoma, a system that represents a small portion of the Ouachita Mountains ecoregion described by Omernik (1987). Within-region heterogeneity will confound attempts to make comparisons among sites, unless the factors responsible for structuring fish assemblages are understood, quantified, and taken into account in the assessments. Typically, water quality agencies in the United States have classified streams and rivers by ecoregion and some surrogate measure of waterbody size. Then, investigators considered the results of physical habitat assessments to determine stream potential (Barbour et al. 1999), although the weights that should be placed on individual habitat attributes remain undetermined for most regions. An alternate approach that may be required in some regions, such as the plains and valleys of Oklahoma, is to use additional factors in the initial classification to further partition variance in fish assemblages and better estimate stream potentials.

In an effort to identify the physical habitat variables that determine stream potential and the structure of fish assemblages of eastern Oklahoma reference streams, I conducted a partial canonical correspondence analysis using CANOCO Windows version 4.0 (ter Braak and Šmilauer 1998). Although similar to the detrended correspondence analysis, the canonical ordination was constrained by linear combinations of

environmental variables, and the detrending step was not required (Palmer 1993). This component of the study was limited to 49 of the reference streams for which data compilation has been completed. The variance attributable to reach volume, estimated from depth and width measurements made along transects, was extracted or "partialled out" to focus on the effects of other environmental variables. Reach volume provided a more direct measure of waterbody size than watershed size, a commonly used surrogate measure.

Fish species abundance data were square root transformed for CCA, as for DCA. I conducted CCA both without transforming environmental variables, and after applying transformations that improved the homoscedasticity and normality of the distributions of individual variables (Table 1). Some of the variables were normally distributed without transformation, and others approached normality after applying a square root transformation. The natural log transformation was used for extremely skewed variables, after adding a constant of 1.0 to each value. Some missing values were encountered for the water quality variables. The missing values were replaced with the median of all values for the variable. The variance inflation factors for each environmental variable were used to check for multicollinearity, as described by ter Braak and Šmilauer (1998).

Table 1. Environmental Variables used in Canonical Correspondence Analysis

Variable Type	Variable	Transformation
Covariable	Reach Volume (m^3)	n/a
Substrate	Silt (%)	Square Root
Composition	Sand (%)	Logarithmic
	Gravel (%)	Logarithmic
	Cobble (%)	n/a
	Boulders (%)	Logarithmic
	Bedrock (%)	Logarithmic

Table 1. Environmental Variables used in Canonical Correspondence Analysis		
Variable Type	Variable	Transformation
	Particulate Organic Matter (%)	Logarithmic
	Hardpack Clay (%)	Logarithmic
Embeddedness	Mean Embeddedness (%)	Square Root
Habitat Type	Maximum Depth (m)	Logarithmic
and Morphology	Riffles (%)	Logarithmic
	Pools (%)	n/a
	Runs (%)	Logarithmic
	Dry (%)	Logarithmic
	Deposition and Scouring (%)	n/a
Cover Types	Large Wood Debris (m^2)	Logarithmic
	Small Wood Debris (m^2)	Logarithmic
	Roots (m^2)	Logarithmic
	Bedrock Ledges (m^2)	Logarithmic
	Submerged Aquatic Vegetation (m^2)	Logarithmic
	Undercut Banks (m^2)	Logarithmic
	Terrestrial Vegetation (m^2)	Logarithmic
Canopy	Mean Canopy Cover (%)	Square Root
Water Chemistry	pH	n/a
	Conductivity (μmhos/cm)	Logarithmic
	Temperature (ºC)	n/a
	Turbidity (NTU)	Logarithmic
	Alkalinity (mg/L)	Square Root
	Dissolved Oxygen (mg/L)	Logarithmic

Data Quality Assessment

An assessment of the quality of available data was conducted to ensure that the results of the study were not compromised (Appendix C). The cluster analysis, described above, served as the primary mechanism for assessment of the quality of fish collections, via comparison of collections made at the same sites in different years. The assessment was based on the assumption that collections made during index periods within the same site over time will be less variable than collections between different sites. Twenty-seven collections from 13 sites were available for the assessment, representing about ¼ of the total number of collections used in the study.

The quality of physical habitat data was assessed by comparing the results of independent assessments of selected reaches by different field crews. I assessed 24 data collection events conducted at 12 sites, by calculating the relative percent difference (RPD) between observations or between key indicators calculated from the data. The RPD is calculated by

$$RPD = \frac{|O_1 - O_2|}{\frac{(O_1 + O_2)}{2}} \times 100$$

where O_1 is the first value in a pair of observations or calculated values, and O_2 is the second value in a pair of observations or calculated values. The RPD values were then summarized by percentiles of differences for each type of observation or indicator (Appendix C).

RESULTS AND DISCUSSION

The Commission has collected fish from 243 sites since 1991. Eighty-two of these were selected as reference sites. Among the 82 reference sites, 13 were sampled in multiple years, yielding 97 reference samples. At the time of this study, physical habitat and water chemistry field data were available for 49 of the reference streams. A total of 103 fish species were collected from the reference streams, out of a possible 175 known species in Oklahoma (Cashner and Matthews 1998).

The high-resolution land use and land cover pattern maps (Riitters et al. 2000) provided a viable alternative to site-specific reconnaissance and physical habitat data. Thirty-three sites were selected, based primarily on this information, although site-

specific information collected by Commission staff proved invaluable in refining the selections. There was a general agreement between the interpretation of potential impairment from land cover and land pattern maps and the field notes and physical habitat characteristics described by Commission staff where both data sets were available. Aside from the variability that appears to be inherent in plains streams, there were no apparent problems with the selected sites with an exception, the tributary of Red Oak Creek.

Significant Clusters

An attempt to determine the number of clusters or groups using statistical tests on cluster analysis was less than successful, due to the continuum of species turnover among sites sampled along spatial gradients. Pillar's (1999) method for determining the numbers of significant clusters indicated that only 3 distinct groups of samples or sites were present, based on analysis of square root transformed counts. The application of other transformations to species counts did not significantly alter the outcome of the test. As described below, the ordination methods consistently revealed additional groups or clusters that were readily interpretable and spatially contiguous that supported unique fish assemblages.

I attributed the difference in numbers of groups identified by Pillar's method and the ordination methods to the continuum of species turnover between adjacent ecoregions and the “fuzziness” introduced by transition zones between regions. Adjacent regions often shared many species (Appendix D), although the species were present in differing relative abundances, both between regions and within regions, dependent on physical habitat conditions at individual sites. Transition zones further blurred the distinction or

partitions between regions, and, as a result, the assignment of sites to groups by cluster analysis changed dependent on the specific samples included in the analysis. I conducted additional cluster analyses and ordinations after removing 1 or 2 samples, and observed a more pronounced effect on the final interpretation of clusters than on groups identified by the ordination methods, described below. In short, the acceptance of "fuzzy partitions" may be required to adequately partition variance among fish assemblages from ecoregions separated by "fuzzy" transition zones.

Detrended Correspondence Analysis

Detrended correspondence analysis of square root transformed fish species counts revealed 6 groups of samples or sites, 2 small outlier groups, and 5 individual samples whose group membership was uncertain (Figure 2). The groups represent (a) the Ozark Highlands, (b) the Boston Mountains, (c) the Ouachita Mountains, (d) the southern part of the Central Oklahoma/Texas Plains, (e) the Flint Hills, Arkansas River Valley, and transition zones around the Ouachita Mountains, and (f) the Central Oklahoma/Texas Plains between the Flint Hills and Red River Valley (Figure 3). The analysis also revealed the fish species that are primarily responsible for determining the positioning of sites in the ordination (Figure 2).

The small outlying groups in the DCA were comprised of faunal rich streams in the Arkansas River basin, and both groups included sites that had been sampled twice in different years. One group included Fourche Maline Creek, Ranch Creek, and Buck Creek. The second group included Blackfork Creek, Greenleaf Creek, and Holson Creek.

Figure 2. Ordination biplot from detrended correspondence analysis of square root transformed fish species counts from wadeable reference streams of eastern Oklahoma. Markers denote site groups and correspond to locations on the following map. The x's denote sites with uncertain group membership. The 25 most important fish species are shown, in positions that depict their influence on the ordination.

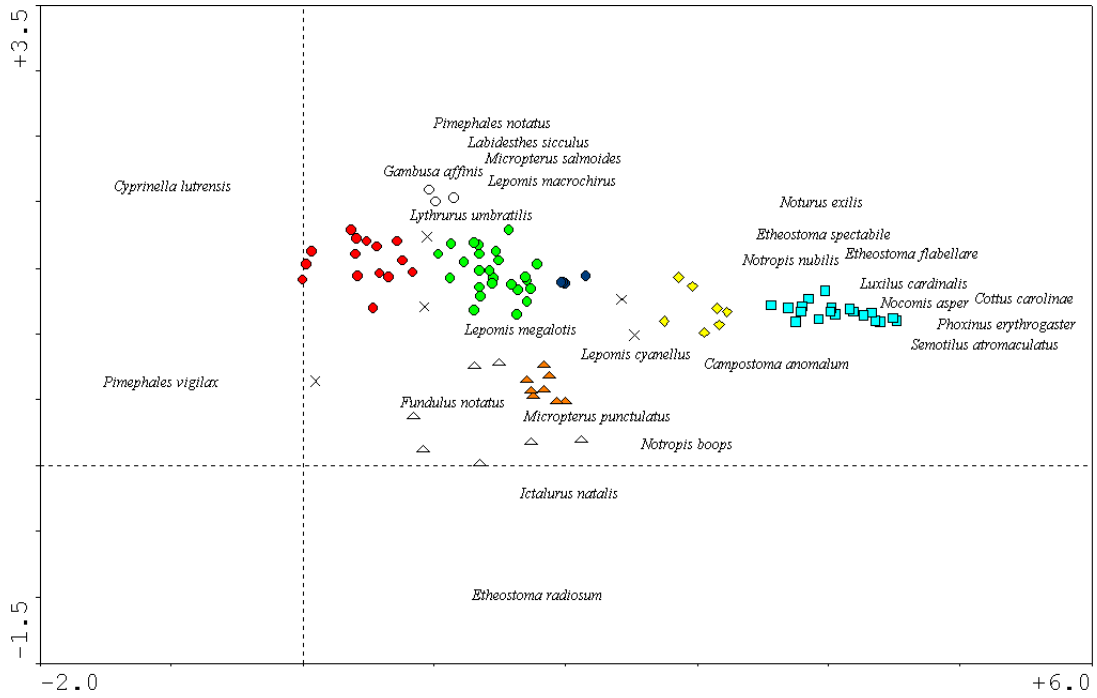
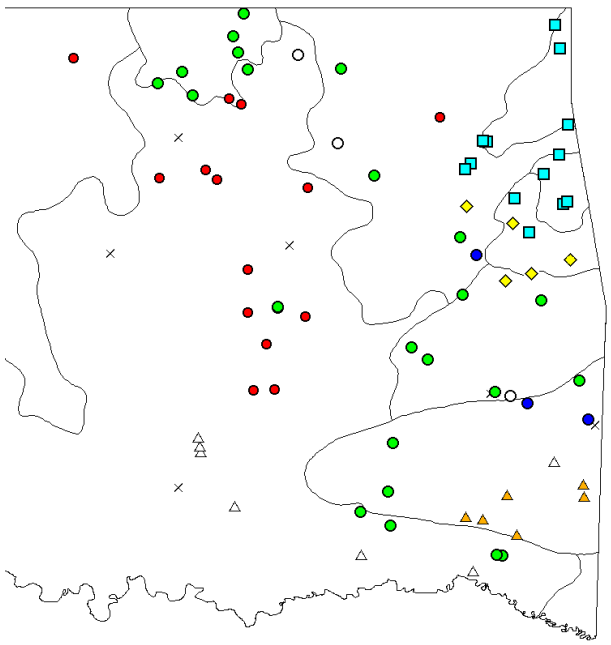


Figure 3. Map of eastern Oklahoma depicting groups identified by detrended correspondence analysis, overlaid by ecoregions described by Omernik (1987).



Nonmetric Multidimensional Scaling

The results of NMDS were initially difficult to interpret, because attempts to depict the ordination in 2 dimensions effectively obliterated distinctions between some of the ecoregions identified by DCA, regardless of the actual number of dimensions specified in the analysis. By reducing the ordination no further than 3 dimensions, and overlaying the results from DCA using different symbol types, sites or samples from unique regions became readily apparent (Figure 4). There was a general agreement between the ordination methods, although the outlying groups identified by DCA were clustered more tightly with other streams in the NMDS ordination, presumably due to the lesser sensitivity of the CYd metric to minor differences in absolute counts. Other differences were also observed in the placement of individual sites in regional groups. A refined map was created to depict the interpretation of the NMDS ordination (Figure 5).

Cluster Analysis

The cluster analysis yielded similar results to the ordination methods (Figure 6). The selected cluster algorithm also identified 2 spatially contiguous groups of sites that may represent distinct regions – the Flint Hills and the transition zones around the Ouachita Mountains (Figure 7) – although, these groups were not readily apparent in the ordinations and they were represented by small numbers of sites. The cluster analysis (Figure 6) also grouped sites in the southern plains with other eastern lowland streams, rather than as a separate cluster. I did not rely heavily on the results of the cluster analysis in the final interpretations, because of the relative instability when compared with the ordinations. The cluster dendogram was helpful in assigning sites to regions where there was some ambiguity in the ordination results. The cluster analysis also provided a useful

Figure 4. Top view (left) and frontal view (right) of ordination of eastern Oklahoma fish assemblages by nonmetric multidimensional scaling on Cao's CYd distance, after reduction to 3 dimensions (stress = 0.11). Symbols depict groups identified by detrended correspondence analysis to show the relative agreement and minor disagreements between the ordinations. Squares and diamonds were removed from the frontal view (right) to avoid obscuring other groups.

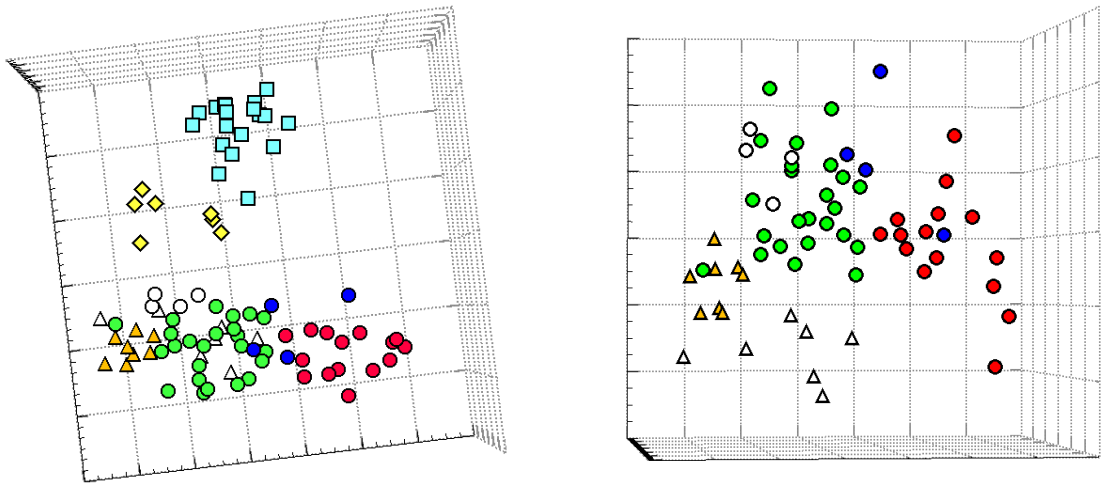


Figure 5. Map of eastern Oklahoma depicting groups of fish assemblages identified by multidimensional scaling, overlaid by ecoregions described by Omernik (1987).

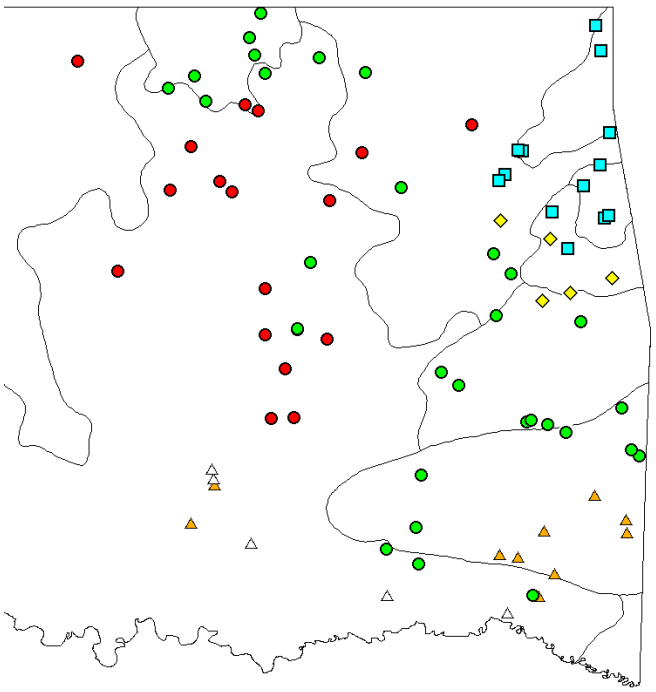


Figure 6. Cluster dendogram produced by grouping reference stream fish assemblages of eastern Oklahoma by Ward's minimum variance algorithm on Cao's CYd distance. Alpha tags denote samples collected from the same waterbody in different years. Symbols on the left correspond to sites on the following map.

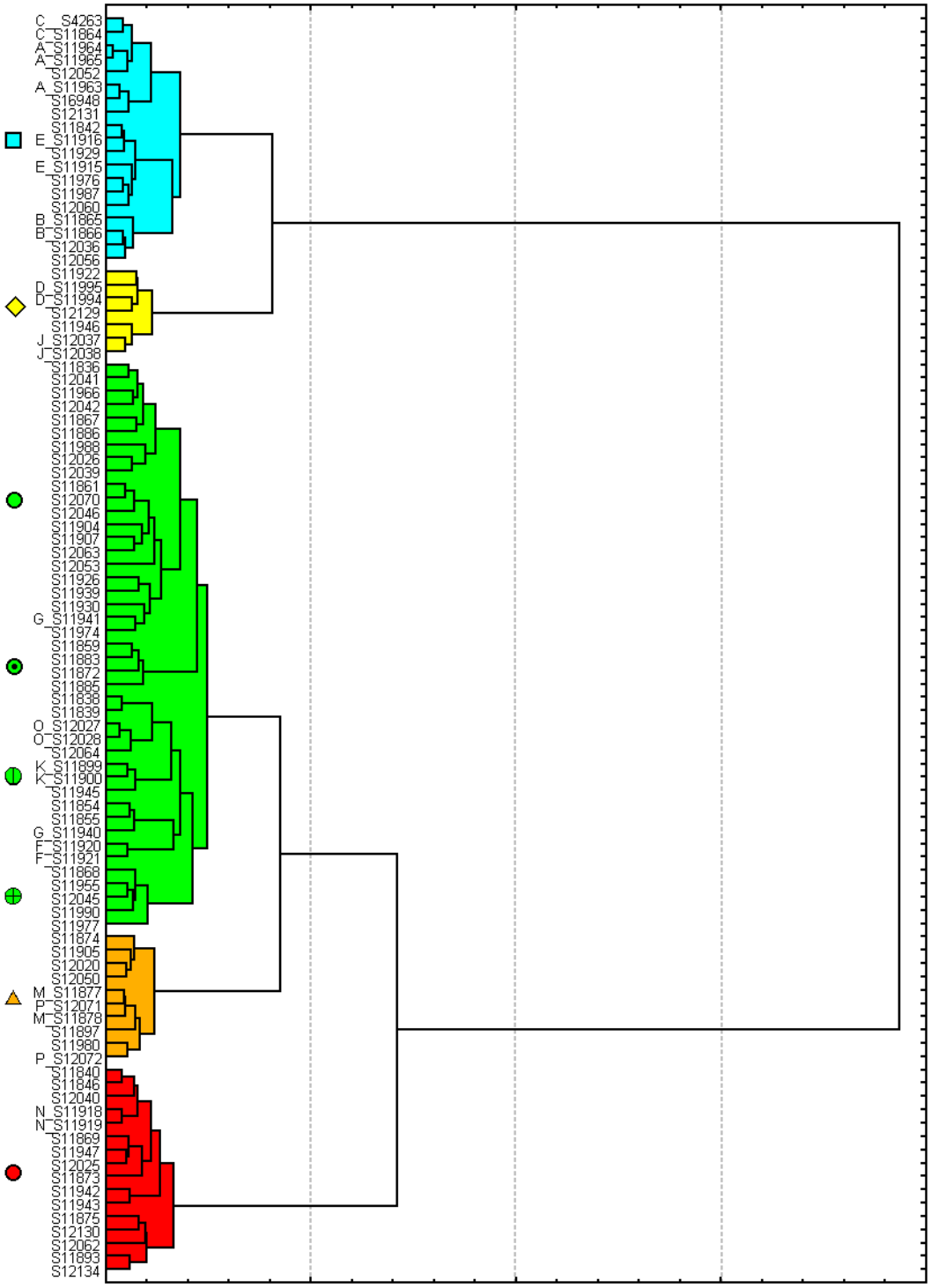
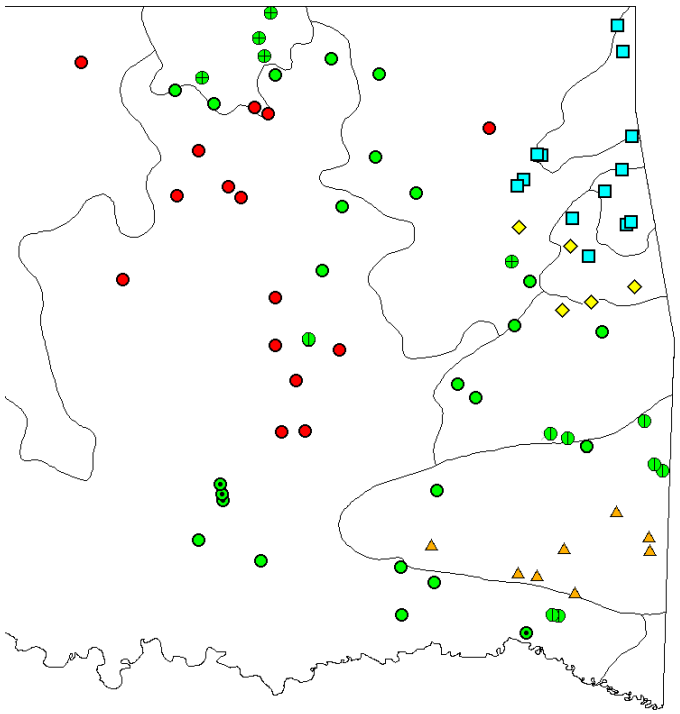


Figure 7. Map of eastern Oklahoma depicting groups of samples or sites identified by cluster analysis on Cao's CYd distance grouped by Ward's minimum variance algorithm. Ecoregions described by Omernik (1987) were overlaid for reference.



mechanism to compare spatial variability with the sum of sampling variability and temporal variability. Samples collected from the same sites in different years clustered as adjacent pairs with few exceptions. Additional details are provided in the data quality assessment (Appendix C).

Summary and Discussion

Matthews and Robison (1988) concluded that the first DCA axis was the best indicator of ecoregions in Arkansas. The results of this study generally agreed with the exception of the Ouachita Mountains and Southern Central Plains, which were split along the second axis from other streams in the broader regions described by Omernik (1987). This split appeared to coincide with the divide between the Red River and Arkansas River basins. Rohm et al. (1987) also observed differences in species composition

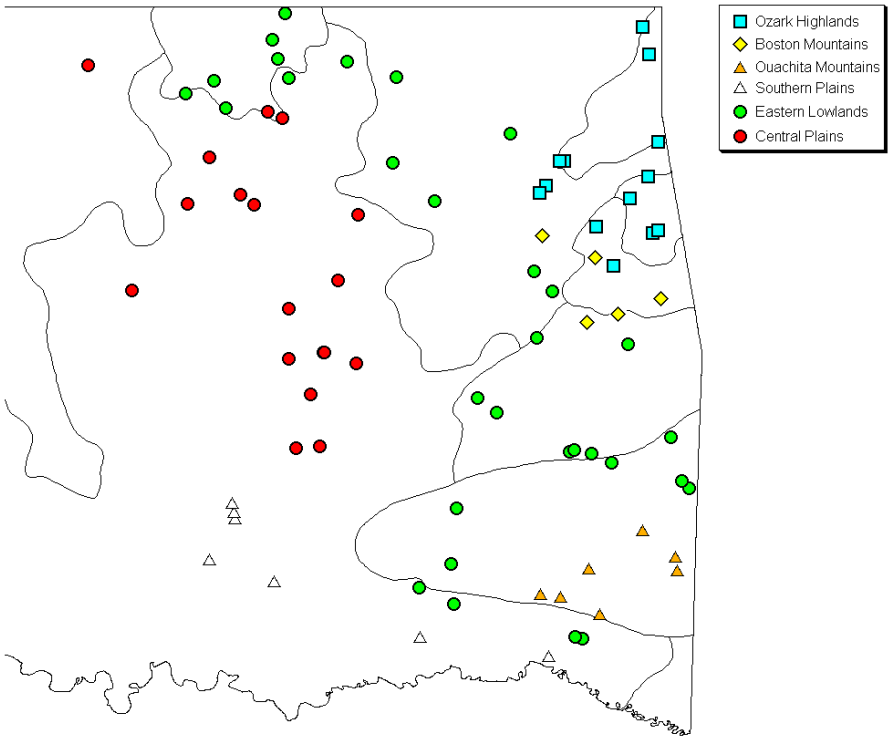
between sites in the Ouachita Mountains ecoregion in Arkansas that coincided with the divide between the Arkansas River and Ouachita River basins. In conclusion, at least part of the "within region heterogeneity" in the Central Oklahoma/Texas Plains ecoregion, described by Toepfer et al. (1988), appears to be attributable to the fact that their study area actually encompasses 2 unique subregions in Oklahoma.

The DCA yielded a more readily interpretable two-dimensional ordination than NMDS. The relatively large number of reference stream collections (n=90) made NMDS more cumbersome, although interpretable results were obtained by reducing the ordination to no fewer than 3 dimensions. There was a general agreement between the DCA and NMDS ordinations with differences in group membership limited to only a few sites.

The potential effects or bias associated with annual variability was insignificant, based on comparison of collections made from individual sites in different years. Although differences in species composition and relative abundance were great enough to separate 1 set of samples collected from the same site over time in the cluster analysis, all same-site pairs grouped closely in the ordinations. The effects of waterbody size were also minimal, as the regions that were identified generally included a wide range of wadeable stream sizes, and the percentage of variance explained by reach volume in the CCA, discussed below, was only 18%.

The results were summarized by producing a final compromised grouping of sites that were more or less spatially contiguous (Figure 8). The table that follows lists the 5 most dominant and 5 most rare species in each region depicted in the map (Figure 8), and the relative abundance of all species are included in Appendix D.

Figure 8. Site groups used for data summary, and comparison of within and between region variability in key metrics of fish community integrity.



The "Eastern Lowlands" is a misnomer, used as a temporary descriptor for this study, because the region included streams in the northern Ouachita Mountains, the Arkansas Valley, the Flint Hills, and the transition areas around the Ouachita Mountains. These streams were not classified into smaller groups, because the combined information from ordination and classification did not yield any readily apparent groups. Further studies are needed to better classify streams within this region, although there were indications in the analyses that the resulting groups will not be spatially contiguous, but rather dependent on specific physical habitat attributes at individual sites, as described by Echelle and Schnell (1976). The streams characterized in this region exhibited the greatest variability, although the region also included the largest number of collections.

Table 2. Dominant and Rare Species in the Redefined Ecoregions					
Ozark Highlands	**Boston Mountains**	**Ouachita Mountains**	**Southern Oklahoma Plains**	**Eastern Lowlands**	**Oklahoma Plains**
Top 5 Dominant Species in each Region					
Luxilus cardinalis	*Campostoma anomalum*	*Lepomis megalotis*	*Campostoma anomalum*	*Campostoma anomalum*	*Cyprinella lutrensis*
Campostoma anomalum	*Luxilus cardinalis*	*Campostoma anomalum*	*Lepomis megalotis*	*Lepomis megalotis*	*Lepomis megalotis*
Phoxinus erythrogaster	*Notropis boops*	*Notropis boops*	*Cyprinella venusta*	*Lepomis cyanellus*	*Pimephales vigilax*
Cottus carolinae	*Notropis nubilus*	*Lepomis cyanellus*	*Etheostoma radiosum*	*Labidesthes sicculus*	*Gambusia affinis*
Semotilus atromaculatus	*Lepomis megalotis*	*Etheostoma radiosum*	*Lepomis cyanellus*	*Gambusia affinis*	*Lepomis cyanellus*
Five Rarest Species in Each Region					
Lepisosteus platostomus	*Minytrema melanops*	*Minytrema melanops*	*Ictalurus melas*	*Amia calva*	*Etheostoma gracile*
Notropis boops	*Lepomis microlophus*	*Notemigonus crysoleucas*	*Percina caprodes*	*Dorosoma pertenense*	*Micropterus dolomieui*
Minytrema melanops	*Notemigonus crysoleucas*	*Micropterus salmoides*	*Cyprinus carpio*	*Pimephales promelas*	*Minytrema melanops*
Labidesthes sicculus	*Percina copelandi*	*Moxostoma duquesnei*	*Anguilla rostrata*	*Carpiodes carpio*	*Etheostoma radiosum*
Lepomis microlophus	*Amboplites ariommus*	*Notropis atrocaudalis*	*Aplodinotus grunniens*	*Noturus gyrinus*	*Notemigonus crysoleucas*

The following figures depict within and between region variability in key measures of fish community integrity, including Cao's CYd distance, species richness, habitat quality tolerance, and water quality tolerance (Figures 9 - 12).

There were an insufficient number of collections in all of the regions to adequately quantify the variance in the selected metrics; however, patterns emerged from the box-and-whisker summaries of key metric values that depict unique attributes of fish communities supported within each region. The metrics generally agree with the results from ordination, and depict minimal overlap between the major groups. The Ozark Highlands assemblages were the most distinct, while the Boston Mountains and Ouachita Mountain assemblages overlapped, apparently because they shared similar substrates and

Figure 9. Within and between ecoregion variability in CYd distance, depicted by box-and-whisker plots. Values were calculated by comparing every collection to a collection from Holson Creek in the Eastern Lowlands (Sample No. 11940).

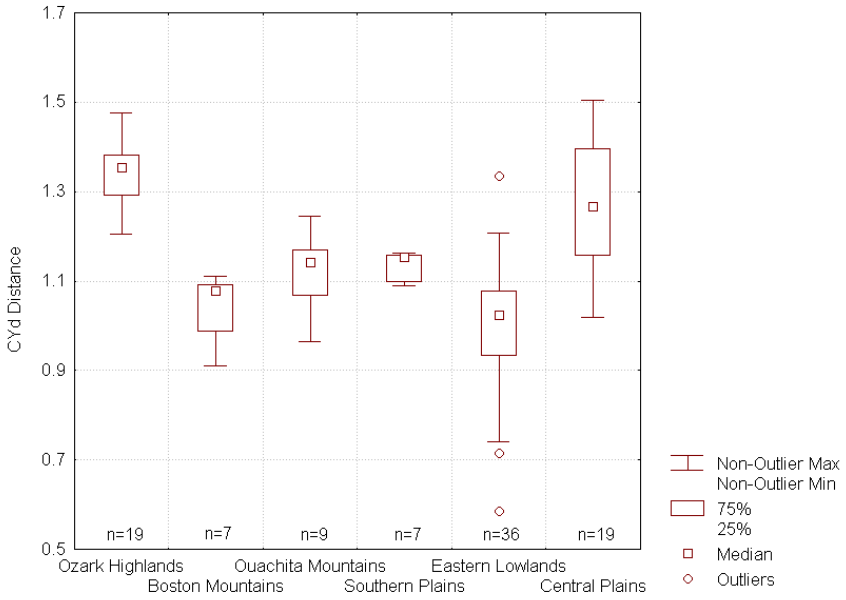


Figure 10. Within and between ecoregion variability in species richness, depicted by box-and-whisker plots.

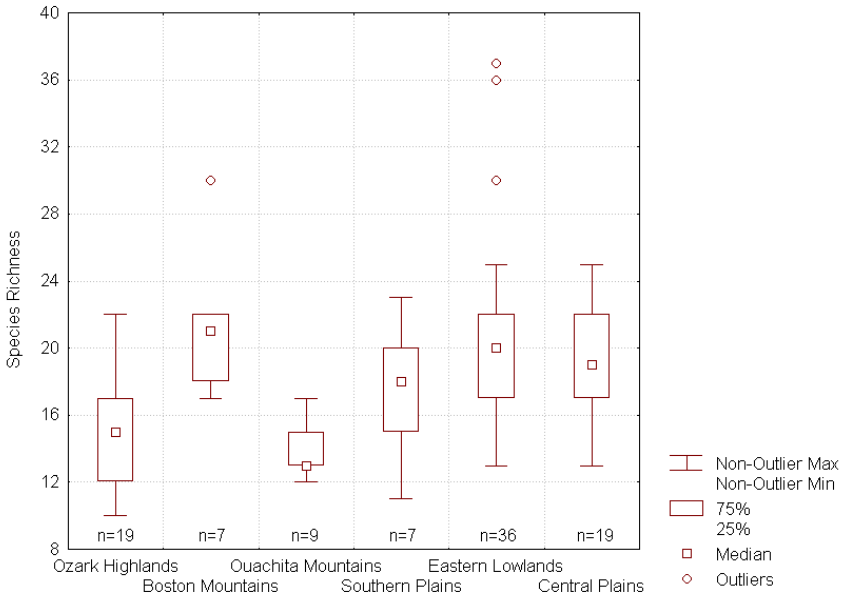


Figure 11. Within and between ecoregion variability in abundance weighted average habitat quality tolerance values (Jester et al. 1992).

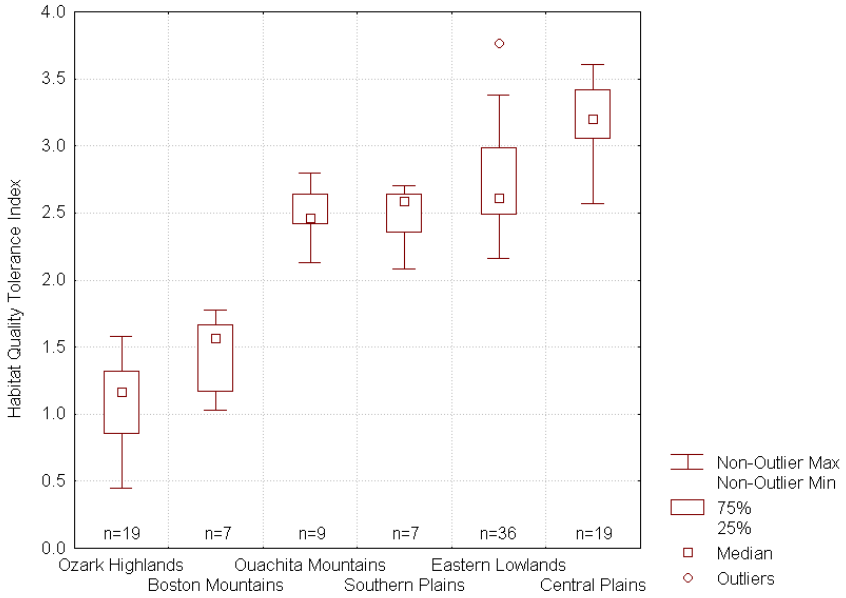
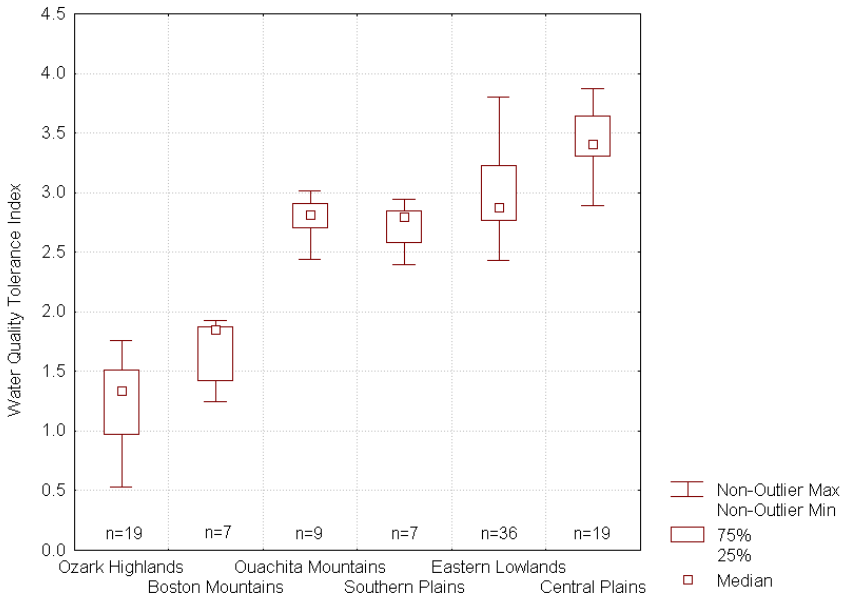


Figure 12. Within and between ecoregion variability in abundance weighted average water quality tolerance values (Jester et al. 1992).



Arkansas River Valley transition zones, at least in the northern part of the Ouachita Mountains. The southern parts of the Ouachita Mountains shared much in common with the Southern Oklahoma Plains region further to the west, perhaps due to their presence within the Red River Basin and/or similarities in substrate composition. Although some species were dominant in multiple regions, none of the regions shared the same dominant species among the top 5.

Canonical Correspondence Analysis

Attempts to refine the site classifications with CCA were unsuccessful, because some of the groups that were identified with the indirect ordination methods were lost in constrained ordinations, most notably the group from the Boston Mountains. A similar result was observed both with and without transformation of environmental variables. It was unclear whether this was a result of a missing environmental variable or variables that induce a shift in species composition, or whether the effects of specific environmental variables overshadowed the effects of subtle differences in species composition and relative abundance, resulting in the dispersion of site groups in the CCA ordination. There was a very small difference between the first eigenvalue in the partial CCA (0.62), using all of the transformed environmental variables, and the first eigenvalue in a partial correspondence analysis (0.65) using the same dataset, minus the environmental variables, indicating that a missing variable was an unlikely cause of the loss of resolution among sites (Palmer 2000).

The CCA ordinations yielded similar results with transformed and untransformed environmental variables, although a lesser degree of multicollinearity was indicated among the transformed variables. The transformations also increased the weight of the

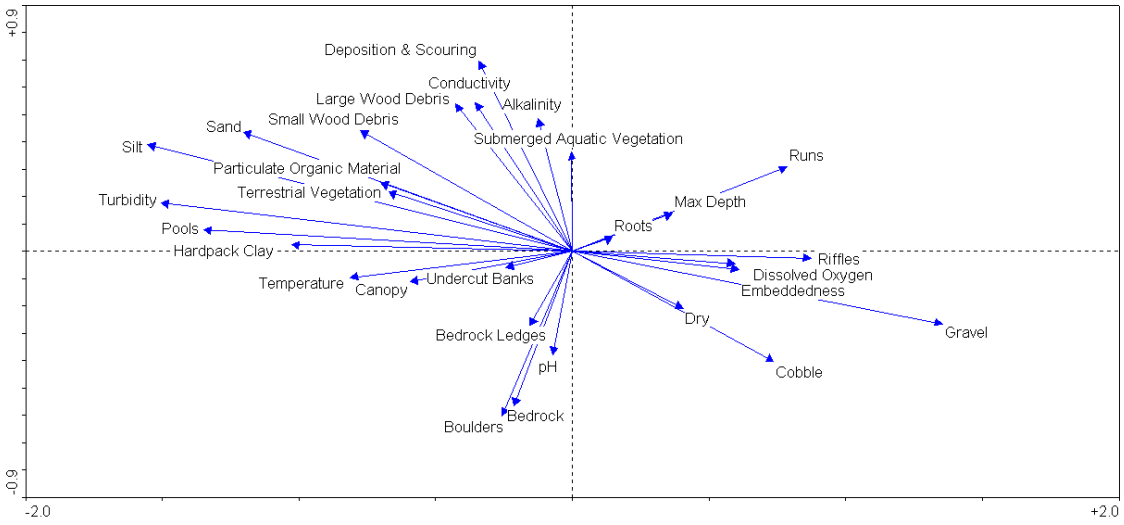
cover type variables on the ordination axes, offering an intuitively appealing improvement. The variance inflation factors for all transformed environmental variables were less than 20; therefore, all 29 of the transformed variables were included in the final analysis. The only elevated inflation factors were values for large woody debris and small woody debris, which were highly correlated (r=0.86). In more refined studies, focused on a single region, it would be prudent to examine the effects of combining these as a single variable.

The total inertia or relative variance in the species abundance matrix produced by the CCA equaled 3.378, the sum of all unconstrained eigenvalues equaled 3.197, and the sum of all canonical eigenvalues equaled 2.294. These values indicated that the reach volumes, as an indicator of waterbody size, explained 18% of the total variance. The transformed environmental variables explained 68% of the remaining variance. I attributed the relatively high percentage of explained variance to the wide variety of streams included in the study area, ranging from high gradient gravel bedded riffle-run streams in the mountains to sandy bottom riffle-pool streams in the plains. I did not attempt to refine a model to assign sites to the regions, because of the wide variety of stream types within the present study area.

The following figure depicts the contribution of selected variables to the structure of reference stream fish assemblages. The lengths and directions of the arrows in the figure depict the relative weight or loading of each variable on the canonical axes. The 49 samples included in CCA appeared to be a representative sample of the 82 sites used in the unconstrained DCA; therefore, the site groups depicted in the DCA ordination

(Figure 2) roughly corresponded to the ordination of environmental variables shown in the following figure.

Figure 13. Relative contributions of 29 selected environmental variables to the structure of reference stream fish assemblages, based on partial canonical correspondence analysis of 49 fish collections from eastern Oklahoma. Variance attributable to stream size, i.e., reach volume, was extracted, before performing the analysis. Transformations were applied to the environmental variables (see Table 1).



Not surprisingly, substrate types and habitat types loaded heavily on the first canonical axis and were the major factors differentiating between fish assemblages of the mountains and plains of eastern Oklahoma. Water quality field measurements loaded on the second axis with the exception of dissolved oxygen, which was correlated with the prevalence of riffles and, therefore, higher gradient. Few of the environmental variables were important in structuring the fish assemblages of the mountains in northeastern Oklahoma, resulting in a strong alignment of samples from these regions along the first canonical axis. These findings appeared consistent with observations made by Giese et al. (1987), who described that the high gradient streams of the Ozark Highlands and Boston Mountains have lesser canopy and high flushing flows that wash away wood debris, in

addition to having more stable substrates. Different stable substrates, bedrock and boulders, loaded heavily on the second canonical axis, and appeared to be the primary factors responsible for distinguishing assemblages in the Ouachita Mountains and Southern Oklahoma Plains from other regions to the north.

The DCA ordination appeared to depict a "tongue effect" or compression on one end of the first ordination axis, a common problem with the detrending and rescaling required to remove the arch effect (Wartenburger 1987). However, the CCA results indicated the configuration of the ordination was attributable to an unbalanced sampling design that included more plains streams than mountain streams, the latter having less variable physical habitat and water quality conditions.

CONCLUSIONS AND RECOMMENDATIONS

There was a general correspondence between most ecoregions described by Omernik (1987) and the structure and composition of fish assemblages of eastern Oklahoma, although there were differences in species composition within regions that may be attributable to the effects of drainage basins. Substrate characteristics and habitat type were the primary factors associated with differences in the composition and structure of fish assemblages.

The results of ordination and classification suggested that the Ouachita Mountains and Central Oklahoma/Texas Plains should be split into subregions, along the northern edge of the Red River basin. Many species were observed in the Arkansas River Basin that were not present within the Red River Basin, although most of these were observed

in low relative abundance. The CCA results suggested that the differences observed between the basins may be attributable to increasing prevalence of boulders and bedrock in the heart of the Ouachita Mountains and Southern Central Plains.

The Ozark Highlands and Boston Mountains regions extended further west than depicted in the available GIS layers, but predictions that these ecoregions would support distinct fish assemblages were accurate. There were some indications that the Flint Hills may also represent a unique ecoregion in Oklahoma, although the streams in this region support fish assemblages that were very similar to others in the Arkansas River Basin. An insufficient number of reference sites were available from the Central Irregular Plains to draw conclusions about the fish assemblages within the region, although they appear to support similar assemblages as others in the Arkansas River basin.

Although these results did not support the decision by Jarman (1984) to combine the eastern Oklahoma regions, he did accurately predict the high variability of habitat conditions and fish assemblages in the plains regions. There remains a clear need for further refinement of the classification, primarily for the Eastern Lowlands and Central Plains streams, because fish assemblages within these regions were highly variable. A more aggressive transformation of fish species abundance may be required to identify groups within these broader regions with DCA. Some preliminary ordinations appeared to indicate that the "square root of the square root" transformation, advocated by Thorne et al. (1999), might yield a higher degree of resolution between site groups in the DCA. In addition, the removal of collections from the Ozark Highlands and Boston Mountains may help to identify smaller groups or classes within the other regions, by removing their influence on the ordinations. The Ouachita Mountains region should be included in the

proposed second phase of this study, because the region supports similar fish assemblages as those in the Southern Oklahoma Plains.

Results from the cluster analysis and a closer scrutiny of ordination results suggested that sites supporting similar assemblages within the Eastern Lowlands and Central Plains regions may not be spatially contiguous, but rather grouped according to site-specific physical habitat conditions. The development of discriminant models may be necessary to adequately partition variance in fish assemblages within these regions, because there is much remaining variability in key bioassessment metrics, even after accounting for stream size.

Appendix A. Visual Basic Code for Calculation of Cao's CYd

```
Option Compare Binary
Option Explicit
    ' The following code was modified from Ludwig and Renolds (1988)
    ' This macro will calculate the CYd measure of (dis)similarity
    ' for every possible combination of samples in a matrix. The process
    ' has not been automated, i.e., it is necessary to manually name the
    ' block of values as "Matrix" on the "Data" worksheet.  Then, manually
    ' enter the number of samples (N) and columns (K) in the first
    ' procedure, below, named "ReadMatrix".
    '
    '         Samples
    ' Species 1   2   3
    ' -----------------
    ' Bass    3   5   5
    ' Carp    0   0   1
    ' Catfish 5   7   3
    ' Sunfish 2   3  12
    '
    Option Base 0
    Dim X As Variant                              'Array for the matrix
    Dim I As Integer, J As Integer                'Counters (X) for the array
    Dim N As Integer, K As Integer, L As Integer 'Number of samples, species, counter
    Dim SC As Integer, TSC As Integer             'Number of species in a pair of samples
    Dim A As Double, B As Double, C As Double     'Used for X(I, J) and X(L, J)
    Dim D As Single                               'CYd Distance
    Dim Cell As Variant
Sub ReadMatrix()
    Set X = Worksheets("Data").Range("Matrix")
        For Each Cell In Worksheets("Data").Range("Matrix").Cells
            If Cell.Value = "" Then Cell.Value = 0.1
        Next                                      'Replace zeros with 0.1
            N = 15                                'N = number of samples
            K = 95                                'K = number of species
        CYd
End Sub
Sub CYd()
  Open "C:\Windows\Desktop\CYdValues.txt" For Output As #1
    For I = 1 To N
        For L = 1 To N
            D = 0
            SC = 0
            TSC = 0
            For J = 1 To K
                If X(J, I) = 0.1 And X(J, L) = 0.1 Then
                    SC = 0                        'Count number of species
                    Else
                    SC = 1
                End If
                A = X(J, I)
                B = X(J, L)
                C = X(J, I) + X(J, L)
                TSC = TSC + SC
                'Debug.Print "A = " & A & " B = " & B
                D = D + (C * Log10(C / 2) - A * Log10(B) - B * Log10(A)) / C
        Next
        D = D * 1 / TSC
        Debug.Print "Sample " & I & " Sample " & L & " CYd = " & D
        Print #1, "Sample " & I & " Sample " & L & " CYd = " & D
        Next
    Next
  Close #1
End Sub
```

Appendix B. List of Reference Streams

Sample	Waterbody ID	Site Name	County
4263	OK121600-05-0140G	Brush Creek	Delaware
11836	OK121500-02-0150G	Adams Creek	Wagoner
11838	OK520500-01-0200U	Alabama Creek (above CC)	Okfuskee
11839	OK520500-01-0200R	Alabama Creek (below CC)	Okfuskee
11840	OK520500-01-0170G	Bad Creek	Okfuskee
11842	OK121700-06-0040G	Battle Creek	Delaware
11846	OKTEMP-0414	Big Creek	Hughes
11854	OK220100-02-0040K	Blackfork Creek	LeFlore
11855	OK220100-02-0040T	Black Fork of Poteau River	LeFlore
11859	OK410400-04-0110G	Bois d'Arc Creek	Pontotoc
11861	OKTEMP-0058	Bolen Creek	Pittsburg
11864	OK121600-05-0140G	Brush Creek	Delaware
11865	OK121600-03-0520G	Brush Creek	Ottawa
11866	OK121600-03-0520G	Brush Creek	Ottawa
11867	OKTEMP-0071	Buck Creek	Osage
11868	OKTEMP-0074	Buck Creek	Osage
11869	OKTEMP-0076	Buckeye Creek	Creek
11872	OK410400-04-0090G	Mill Creek	Pontotoc
11873	OKTEMP-0102	Camp Creek	Pawnee
11874	OK410210-02-0240G	Caney Creek	Pushmataha
11875	OKTEMP-0111	Captain Creek	Lincoln
11877	OK410210-08-0120G	Cedar Creek	McCurtain
11878	OK410210-08-0120G	Cedar Creek	McCurtain
11883	OK410400-04-0010T	Clear Boggy Creek	Pontotoc
11885	OKTEMP-0133	Clear Creek	McCurtain
11886	OKTEMP-0136	Clear Creek	Osage
11893	OKTEMP-0156	Council Creek	Payne
11897	OK410210-06-0210G	Cucumber Creek	LeFlore
11899	OK410210-01-0070G	Cypress Creek	McCurtain
11900	OK410210-01-0070G	Cypress Creek	McCurtain
11904	OKTEMP-0171	Doga Creek	Osage
11905	OK410210-06-0270G	Dry Creek	McCurtain
11907	OK410300-03-0210C	Dumpling Creek	Pushmataha
11915	OK121700-05-0140G	England Creek	Adair
11916	OK121700-05-0140G	England Creek	Adair
11918	OK520500-01-0280G	Flat Rock Creek	Okfuskee
11919	OK520500-01-0280G	Flat Rock Creek	Okfuskee
11920	OK220100-04-0010M	Fourche Maline Creek	LeFlore
11921	OK220100-04-0010M	Fourche Maline Creek	LeFlore
11922	OKTEMP-0195	Fourteen Mile Creek	Cherokee
11926	OKTEMP-0210	Gray Horse Creek	Osage
11929	OK121700-05-0150G	Green Creek	Adair
11930	OKTEMP-0211	Greenleaf Creek	Muskogee

Sample	Waterbody ID	Site Name	County
11939	OKTEMP-0222	Hogshooter	Washington
11940	OK220100-04-0030G	Holson Creek	LeFlore
11941	OK220100-04-0030G	Holson Creek	LeFlore
11942	OKTEMP-0225	Hominy Creek	Osage
11943	OKTEMP-0227	Hominy Creek	Osage
11945	OK410210-01-0060G	Horsehead Creek	McCurtain
11946	OK220200-05-0050G	Jenkins Creek	Adair
11947	OKTEMP-0237	Lagoon Creek	Creek
11955	OKTEMP-0241	Little Chief Creek	Osage
11963	OK121600-02-0070G	Little Saline Creek	Mayes
11964	OK121600-02-0070G	Little Saline Creek	Mayes
11965	OK121600-02-0070G	Little Saline Creek	Mayes
11966	OK220200-02-0040G	Little Sallisaw Creek	Sequoyah
11974	OKTEMP-0259	Longtown Creek	Pittsburg
11976	OK121700-03-0260G	Luna Creek	Adair
11977	OKTEMP-0007	Bayou Manard	Muskogee
11980	OK410210-06-0060G	Mine Cr	McCurtain
11987	OK121700-04-0020G	Negro Jake Creek	Cherokee
11988	OK120420-03-0040G	Nickel Creek: 91st Street	Tulsa
11990	OKTEMP-0272	N. Bird Creek	Osage
11994	OK121700-02-0270G	Park Hill Creek	Cherokee
11995	OK121700-02-0270G	Park Hill Creek	Cherokee
12020	OKTEMP-0301	Pine Creek	Pushmataha
12025	OKTEMP-0312	Pryor Creek	Mayes
12026	OK121300-01-0060G	Ranch Creek	Tulsa
12027	OK220100-04-0050G	Red Oak Creek (Downstream-Below	Latimer
12028	OK220100-04-0050G	Red Oak Creek (Downstream-Below	Latimer
12036	OKTEMP-0335	Saline Creek	Mayes
12037	OK220200-03-0010G	Sallisaw Creek	Sequoyah
12038	OK220200-03-0010G	Sallisaw Creek	Sequoyah
12039	OK520700-02-0150G	Salt Creek	Okmulgee
12040	OKTEMP-0415	Salt Cr	Hughes
12041	OKTEMP-0333	South Fork Dirty Creek	Muskogee
12042	OKTEMP-0346	San Bois Creek	Pittsburg
12045	OKTEMP-0354	Sand Creek	Osage
12046	OK410400-03-0160T	Sandy Creek	Johnston
12050	OK410210-09-0100G	Silver Cr	McCurtain
12052	OKTEMP-0368	Snake Creek	Mayes
12053	OK310800-01-0160G	Spring Creek	Johnston
12056	OKTEMP-0374	Spring Creek	Mayes
12060	OK121700-03-0120G	Steely Hollow	Cherokee
12062	OKTEMP-0381	Stink Creek	Kay
12063	OK410400-01-0200G	Sugar Creek	Choctaw
12064	OK220100-01-0160G	Sugar Loaf Creek	LeFlore
12070	OKTEMP-0385	Tenmile Creek	Pushmataha

Sample	Waterbody ID	Site Name	County
12071	OK410210-02-0150G	Terrapin Creek	Pushmataha
12072	OK410210-02-0150G	Terrapin Creek	Pushmataha
12076	OK220100-04-0050-08N	Unnamed Trib. to Red Oak (Oak	Latimer
12077	OK220100-04-0050-08N	Unnamed Trib. to Red Oak (Oak	Latimer
12129	OK220200-02-0130G	Vian Creek	Sequoyah
12130	OK520700-03-0020G	Walnut Creek	Okfuskee
12131	OK121600-07-0050G	Warren Branch Creek	Ottawa
12134	OKTEMP-0395	Wewoka Creek	Hughes
16948	OK121600-02-0070F	Little Saline Creek	Mayes

Appendix C. Data Quality Assessment

The Oklahoma Conservation Commission maintains and periodically updates a set of standard operating procedures that describe their methods for data collection (OCC 1996). New employees are provided with a copy of the procedures, and trained in their use by experienced field staff. The more experienced field staff periodically conduct technical audits of data collection procedures practiced by those responsible for routine data collection.

Field notes document sample collection efforts and any problematic conditions that affect the integrity or representativeness of individual samples. For example, fish collection records include fish species that were observed, but not collected, the amount of time spent shocking, along with voltage and amperage settings, and the amount of time spent seining, along with seine dimensions. Also, fish collection records include notes about factors that may affect collection efficiency, such as the effects of low conductivity on electrofishing success and the effects of cobble substrate on seining success.

One data quality issue that limited some aspects of the study was the incompleteness of data, resulting from the immature status of efforts to compile the data. For example, although excellent records are maintained about fish data collection efforts, these data have been compiled for only about 2/3 of the reference streams. Completeness was better for those variables for which data entry has been completed. For example, most water quality variables were 100% complete, although pH records were about 80% complete. Velocity and discharge data were available for only about half of the streams, although these variables would probably be correlated with the habitat type and morphology variables used in the study.

Two fish species were included in the database that appear to be the same species, i.e., *Luxilus cardinalis* and *Notropis (Luxilus) pilsbryi*. Cashner and Matthews (1988) describe that *Luxilus pilsbryi* is restricted to the White River drainage in Missouri and Arkansas. Therefore, for the purposes of my analyses, I assumed all individuals were *L. cardinalis*.

Replicate fish collections are difficult to obtain (Fore et al. 1994); however, we may generally assume that collections made during index periods within the same site over time will be less variable than collections between different sites. By comparing similarity between same-site collections, we may assess whether sampling variability is excessive. As depicted in the cluster dendogram (Figure 6), a notable difference in same-site collections was observed between only 1 pair out of a total of 13 same-site pairs. Collections made from Holson Creek, numbered 11940 and 11941, were clustered into different groups, although both collections were clustered within the same major group. The differences between samples 11940 and 11941 did not influence the interpretation of either of the ordinations, because the samples in question were grouped closely together in both DCA and NMDS (depicted by white circles in Figures 2 and 4). In summary, the quality or representativeness of fish collections was adequate to meet the objectives of this study.

Commission staff periodically conducts quality control sampling events to document the quality of physical habitat data. Multiple teams assess the same reach at different times, usually on consecutive days. I compared the data collected from 24 sampling events at 12 sites by calculating the relative percent difference (RPD) between

observations or between key indicators calculated from the data. The RPD is calculated by

$$RPD = \frac{|O_1 - O_2|}{\frac{(O_1 + O_2)}{2}} \times 100$$

where O_1 is the first value in a pair of observations or calculated values, and O_2 is the second value in a pair of observations or calculated values. Extreme RPD values, summarized in tables below, were generally observed only where field staff had to estimate indicators in very small magnitudes. For example, the first team recorded no gravel substrate, whereas the second team recorded 1 percent gravel substrate. I did not observe any differences that would materially alter the conclusions of the study.

Percentiles of relative percent differences (RPD) between pairs of observations or calculated values used in the study.

	Canopy and Waterbody Size			Habitat Types		
Percentile	Average Canopy	Reach Volume	Maximum Depth	Pools (%)	Riffles (%)	Run (%)
5th	2%	0%	0%	0%	0%	0%
25th	5%	1%	0%	5%	0%	9%
Median	7%	2%	0%	11%	0%	11%
75th	14%	14%	22%	17%	0%	15%
95th	103%	28%	42%	27%	118%	115%

	Substrate Types						
Percentile	Bedrock	Boulders	Cobble	Gravel	Hardpack	Sand	Silt
5th	0%	0%	0%	0%	0%	2%	5%
25th	0%	0%	0%	10%	8%	4%	14%
Median	0%	0%	3%	26%	28%	7%	34%
75th	3%	15%	30%	61%	61%	20%	35%
95th	102%	148%	122%	200%	130%	41%	88%

Percentiles of relative percent differences (RPD) between pairs of observations or calculated values used in the study.								
	Cover Types							
Percentile	Bedrock Ledges	Large Woody Debris	Particulate Organic Matter	Roots	Small Woody Debris	Submerged Aquatic Vegetation	Submerged Terrestrial Vegetation	Undercut Banks
5th	1%	4%	13%	7%	2%	0%	5%	0%
25th	51%	6%	21%	13%	8%	0%	15%	33%
Median	93%	11%	34%	17%	16%	74%	33%	50%
75th	200%	29%	51%	38%	24%	200%	104%	93%
95th	200%	85%	130%	80%	53%	200%	168%	116%

Appendix D. Relative Abundance of Fish Species within the Ecoregions

Family	Species	Ozark Highlands	Boston Mountains	Ouachita Mountains	Southern Oklahoma Plains	Eastern Lowlands	Oklahoma Plains
Petromyzontidae	*Ichthyomyzon gagei*					<1	
Lepisosteidae	*Lepisosteus oculatus*					<1	<1
	Lepisosteus osseus					<1	<1
	Lepisosteus platostomus	<1					<1
Amiidae	*Amia calva*					<1	
Anguillidae	*Anguilla rostrata*				<1		
Clupeidae	*Dorosoma cepedianum*					1	2
	Dorosoma petenense					<1	<1
Esocidae	*Esox americanus*			1		<1	
Cyprinidae	*Campostoma anomalum*	2	40	2	3	2	2
	Cyprinella camura					<1	
	Cyprinella lutrensis				2	4	28
	Cyprinella venusta				15	<1	
	Cyprinella whipplei		<1			1	
	Cyprinus carpio		<1		<1	<1	<1
	Hybognathus placitus					<1	<1
	Luxilus chrysocephalus					<1	
	Luxilus cardinalis	42	24			<1	
	Lythrurus fumeus			3	2	<1	
	Lythrurus snelsoni			4		<1	
	Lythrurus umbratilis			<1	2	4	<1
	Nocomis asper	<1	<1				
	Notemigonus crysoleucas		<1	<1		1	<1
	Notropis amblops	<1					
	Notropis amnis				<1		
	Notropis atherinoides				<1	1	1
	Notropis atrocaudalis			<1			
	Notropis boops	<1	14	19	8	<1	
	Notropis emiliae					<1	
	Notropis greenei		<1				
	Notropis nubilus	3	<1			<1	
	Notropis ortenbergeri					<1	
	Notropis ozarcanus		<1				
	Notropis rubellus	<1	<1				
	Notropis stramineus				<1		4
	Notropis volucellus					<1	
	Phenacobius mirabilis				<1	<1	<1
	Phoxinus erythrogaster	22			1		
	Pimephales notatus		<1	2	<1	8	4
	Pimephales promelas					<1	<1
	Pimephales tenellus					<1	<1
	Pimephales vigilax				3	<1	8
	Semotilus atromaculatus	<1	<1	<1			
Catostomidae	*Carpiodes carpio*					<1	<1
	Catostomus commersoni	<1					

Family	Species	Ozark Highlands	Boston Mountains	Ouachita Mountains	Southern Oklahoma Plains	Eastern Lowlands	Oklahoma Plains
	Erimyzon oblongus			2		<1	
	Hypentelium nigricans	<1	<1			<1	
	Ictiobus bubalus					<1	<1
	Minytrema melanops	<1	<1	<1		<1	<1
	Moxostoma duquesnei	<1	<1	<1		<1	
	Moxostoma erythrurum	<1	<1	<1	6	<1	<1
Ictaluridae	*Ictalurus melas*	<1	<1	<1	<1	<1	<1
	Ictalurus natalis	<1	<1	<1	5	<1	1
	Ictalurus punctatus		<1		<1	<1	<1
	Noturus eleutherus			<1			
	Noturus exilis	5	3	<1		1	
	Noturus gyrinus					<1	
	Noturus miurus					<1	
	Noturus nocturnus			<1	<1	<1	<1
	Pylodictis olivaris			<1		<1	<1
Aphredoderidae	*Aphredoderus sayanus*					<1	
Fundulidae	*Fundulus catenatus*	<1					
	Fundulus notatus			1	2	3	<1
	Fundulus olivaceus		<1			2	
	Fundulus zebrinus						<1
Poeciliidae	*Gambusia affinis*	<1	<1		2	10	8
Atherinidae	*Labidesthes sicculus*	<1	<1	6	<1	14	3
Cottidae	*Cottus carolinae*	12	<1				
Percichthyidae	*Morone chrysops*					<1	<1
Centrarchidae	*Amboplites ariommus*	<1	<1				
	Ambloplites rupestris	<1					
	Lepomis cyanellus	<1	4	19	9	16	6
	Lepomis gulosus	<1	<1		<1	2	<1
	Lepomis humilis					1	<1
	Lepomis macrochirus	<1	<1	<1	3	6	<1
	Lepomis marginatus					<1	
	Lepomis megalotis	<1	<1	31	18	2	23
	Lepomis microlophus	<1	<1	<1	<1	<1	<1
	Lepomis punctatus				<1	<1	
	Micropterus dolomieui	<1	<1	<1		<1	<1
	Micropterus punctulatus	<1	<1	<1	1	2	<1
	Micropterus salmoides	<1	<1	<1	<1	2	1
	Pomoxis annularis				<1	<1	<1
	Pomoxis nigromaculatus					<1	<1
Percidae	*Etheostoma asprigene*					<1	
	Etheostoma blennioides	<1	<1			<1	
	Etheostoma chlorosomum					<1	<1
	Etheostoma flabellare	7	2			<1	
	Etheostoma gracile					<1	<1
	Etheostoma histrio					<1	
	Etheostoma nigrum			<1			
	Etheostoma punctulatum	<1	<1			<1	
	Etheostoma radiosum			7	1	<1	<1

Family	Species	Ozark Highlands	Boston Mountains	Ouachita Mountains	Southern Oklahoma Plains	Eastern Lowlands	Oklahoma Plains
	Etheostoma spectabile	<1	3		7	2	<1
	Etheostoma whipplei		<1			1	<1
	Etheostoma zonale					<1	
	Percina caprodes	<1	<1	<1	<1	<1	<1
	Percina copelandi		<1			<1	
	Percina maculata			<1		<1	
	Percina phoxocephala						<1
	Percina sciera				<1	<1	
	Percina shumardi					<1	
Sciaenidae	*Aplodinotus grunniens*				<1	<1	<1

REFERENCES

Anderson AA, Hubbs C, Winemiller KO, Edwards RJ. 1995. Texas freshwater fish assemblages following three decades of environmental change. Southwest Nat 40(3):314-321.

Bailey RG. 1982. Classification systems for habitat and ecosystems, p. 16-26. In Mason WT, Jr. (ed.), Research on fish and wildlife habitat. U.S. Environmental Protection Agency Office of Research and Development, EPA-600-8-82-022.

Barbour MT, Gerritsen J, Snyder BD, Stribling JB. 1999. Rapid bioassessment protocols for use in streams and rivers: Periphyton, benthic macroinvertebrates and fish, 2nd ed. U.S. Environmental Protection Agency Office of Water, EPA 841-B-99-002.

Barbour, MT, Gerritsen J, Griffith GE, Frydenborg R, McCarron E, White JS, Bastian ML. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. J North Am Benthol Soc 15(2):185-211.

Boyle TP, Smillie GM, Anderson JC, Beeson DR. 1990. A sensitivity analysis of nine diversity and seven similarity indices. J Water Poll Control Fed. 62(6):749-762.

Cao Y, Williams WP, Bark AW. 1997a. Similarity measure bias in river benthic Aufwuchs community analysis. Water Env Res 69(1):95-106.

Cao Y, Bark AW, Williams WP. 1997b. A comparison of clustering methods for river benthic community analysis. Hydrobiol 347:25-40.

Cao Y, Williams DD, and Williams NE. 1998. How important are rare species in aquatic community ecology and bioassessment? Limnol Ocean 43(7):1403-1409.

Conquest LL, Ralph SC, Naiman RJ. 1993. Implementation of large-scale stream monitoring efforts: Sampling design and data analysis issues, p. 69–90. In Loeb S and Spacie A. (ed.), Biological monitoring of aquatic systems. Lewis Press, Boca Raton, FL.

Cashner RC, Mathews WJ. 1988. Changes in the known Oklahoma fish fauna from 1973 to 1988. Proc Oklahoma Acad Sci 68:1-7.

Dewalt RE. 1995. Biological communities of reference streams in the South Central Plains and Upper Mississippi Alluvial Plains ecoregions of Louisiana. Prepared for the Louisiana Department of Environmental Quality. Baton Rouge, LA.

Echelle AA, Schnell GD. 1976. Factor analysis of species associations among fishes of the Kiamichi River, Oklahoma. Trans Am Fish Soc 105:17-31.

Efron B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap, and other methods. Biometrika 68: 589-599.

Frenzel SA, Swanson RB. 1996. Relations of fish community composition to environmental variables in streams of central Nebraska, USA. Env Mgmt 20(5):689-705.

Fore LS, Karr JR, Conquest LL. 1994. Statistical properties of an index of biological integrity used to evaluate water resources. Can J Fish Aquat Sci 51:1077-1087.

Frissell CA, Liss WJ, Warren CE, Hurley MD. 1986. A hierarchical framework for stream classification: Viewing streams in a watershed context. Env Mgmt 10(2)199-214.

Gallant AL, Whittier TR, Larsen DP, Omernik JM, Hughes RM. 1989. Regionalization as a tool for managing environmental resources. U.S. Environmental Protection Agency Office of Research and Development, EPA-600-3-89-060.

Giese J, Keith B, Maner M, McDaniel R, Singleton B. 1987. Physical, chemical, and biological characteristics of least disturbed reference streams in Arkansas' ecoregions. State of Arkansas Department of Pollution Control and Ecology. Little Rock, AR.

Grumbine RE. 1994. What is ecosystem management? Cons Biol 8(1):27-38.

Hawkes CL, Miller DL, Layher WG. 1986. Fish ecoregions of Kansas: Stream fish assemblage patterns and associated environmental correlates. Env Biol Fish. 17(4):267-279.

Hirst SM 1984. Applied ecology and the real world: Resource management and impact assessment. J Env Mgmt 18:203-213.

Hilsenhoff WL. 1982. Using a biotic index to evaluate water quality in streams. Wisconsin Department of Natural Resources. Madison, Wisconsin, USA. Technical Bulletin No. 132.

Hornig CE, Bayer CW, Twidwell SR, Davis JR, Kleinsasser RJ, Linam GW, Mayes KB. 1994. Development of regionally based biological criteria in Texas, p. 145-152. In Davis WS, Simon TP (ed.), Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers. Boca Raton, Florida, USA.

Hughes RM. 1984. Use of watershed characteristics to select control streams for estimating effects of metal mining wastes on extensively disturbed streams. Env Mgmt 9(3):253-262.

Hughes RM, Larsen DP. 1988. Ecoregions: An approach to surface water protection. J Water Poll Control Fed 60:486-493.

Hughes RM, Larsen DP, Omernik JM. 1986. Regional reference sites: A method for assessing stream potentials. Env Mgmt 10(5):629-635.

Hughes RM, Heskary SA, Matthews WJ, Yoder CO. 1993. Use of ecoregions in biological monitoring, p. 125-151. In Loeb S. and Spacie A. (ed.), Biological monitoring of aquatic systems. Lewis Press, Boca Raton, FL.

Hughes RM. 1994. Defining acceptable biological status by comparing with reference conditions, p. 31-47. In Davis WS, Simon T.P. (ed.), Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers. Boca Raton, Florida, USA.

Jackson DA. 1997. Compositional data in community ecology: The paradigm or peril of proportions? Ecology 78(3)929-940.

Jarman R. 1984. The development of aquatic ecoregions in Oklahoma [dissertation]. The University of Oklahoma Library. Norman, Oklahoma, USA. 196 p.

Jester DB, Echelle AA, Matthews WJ, Pigg J, Scott CM, Collins KD. 1992. The fishes of Oklahoma, their gross habitats, and their tolerance of degradation in water quality and habitat. Proc Oklahoma Acad Sci 72:7-19.

Jongman RHG, ter Braak CJF, and Van Tongeren OFR. 1995. Data analysis in community and landscape ecology. Cambridge University Press. New York, New York, USA. 299 p.

Kachigan SK. 1991. Multivariate statistical analysis: A conceptual introduction, 2nd ed. Radius Press, New York, New York, USA. 303 p.

Karr JR, Dudley DR. 1981. Ecological perspective on water quality goals. Environmental Management 5:55-68.

Karr JR, Fausch KD, Angermeier PL, Yant PR, Schlosser IJ. 1986. Assessing biological integrity in running waters: A method and its rationale. Illinois Natural History Survey, Urbana, Illinois, USA. Special Publication 5.

Karr JR. 1993. Biological monitoring: Challenges for the future, p. 357-373. In Loeb S and Spacie A (ed.), Biological monitoring of aquatic systems. Lewis Press, Boca Raton, FL.

Karr JR, Chu EW. 1997. Biological monitoring: Essential foundation for ecological risk management. Human Eco Risk Ass 3(6):993-1004.

Karr JR, Chu EW. 1999. Restoring life in running waters: Better biological monitoring. Island Press. Washington D.C. USA. 206 p.

Kenkel NC, Orlóci L. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: Some new results. Ecology 67(4):919-928.

Kovach Computing Services. 2000. Multivariate Statistical Package version 3.12b [Software program] Pentraeth, Anglesey, Wales, United Kingdom.

Legendre L, Legendre P. 1998. Numerical ecology, 2nd English ed. Elsevier, Amsterdam, The Netherlands. 853 p.

Lyons J. 1989. Correspondence between the distribution of fish assemblages in Wisconsin streams and Omernik's ecoregions. Am Midland Nat 122:163-182.

Ludwig JA, Reynolds JF. 1988. Statistical ecology: A primer on methods and computing. John Wiley and Sons. New York, New York, USA. 306 p.

Matthews WJ and Robison HW. 1988. The distribution of fishes of Arkansas: a multivariate analysis. Copeia 1988:358-374.

Matthews WJ, Hough DJ, and Robison HW. 1992. Similarities in fish distribution and water quality patterns in streams of Arkansas: Congruence of multivariate analyses. Copeia 1992(2):296-305.

Maxted JR. Biology, probability, and the obvious. Human Eco Risk Mgmt 3(6):955-965.

McNab WH, Browning SA, Simon SA, Fouts PE. 1999. An unconventional approach to ecosystem unit classification in western North Carolina, USA. Forest Eco Mgmt 114(1999):405-420.

Milligan GW, Cooper MC. 1985. An examination of procedures for determining the number of clusters in a dataset. Psychometrika 50(2):159-179.

Mount D. 1994. A comparison of the strengths and limitations of chemical specific criteria, whole effluent toxicity testing, and biosurveys. Prepared for Science Applications International Corporation, Submitted to Environmental Protection Agency, Office of Wastewater Enforcement and Compliance. Washington D.C.

Nemec AFL, Brinkhurst RO. 1988. Using the bootstrap to assess statistical significance in the cluster analysis of species abundance data. Can J Fish Aquat Sci 45:965-970.

OCC 1996. Oklahoma Conservation Commission standard operating procedures for the collection and analysis of water quality samples. Oklahoma Conservation Commission. Oklahoma City , Oklahoma, USA.

Ohio EPA. 1987. Biological criteria for the protection of aquatic life: Volume I. The role of biological data in water quality assessment. Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio, USA.

Omernik JM. 1987. Ecoregions of the conterminous United States. Ann Assoc Am Geo 77(1):118-125.

Omernik JM, Griffith GE. 1991. Ecological regions versus hydrologic units: Frameworks for managing water quality. J Soil Water Cons 46(5):334-340.

Palmer MW. 1993. Putting things in even better order: The advantages of canonical correspondence analysis. Ecol 74(8):2215-2230.

Palmer MW. 1998. Putting things in order: The advantages of detrended correspondence analysis. Am Nat 131(6):924-934.

Palmer MW. 2000. Ordination methods for ecologists. Oklahoma State University, Stillwater, Oklahoma, USA. <http://www.okstate.edu/artsci/botany/ordinate/> Accessed 2000 Sep 27.

Petersen JC. 1998. Water-quality assessment of the Ozark plateaus study unit, Arkansas, Kansas, Missouri, and Oklahoma: Fish communities in streams of the Ozark Plateaus and their relationships to selected environmental factors. U.S. Geological Survey. Denver, Colorado, USA. Water-Resources Investigations Report 98-4155.

Pillar VD. 1999. How sharp are classifications? Ecology 80(8):2508-2516. <http://esa.sdsc.edu/Archive/E080-014/main.html> Accessed 2000 Oct 5.

Polls I. How people in the regulated community view biological integrity. J North Am Benthol Soc 13(4):598-604.

Richards C, Host GE. 1993. Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. Freshwater Biology 29:285-294.

Riitters KH, Wickham JD, Vogelmann JE, Jones KB. 2000. National land-cover pattern data. Ecological Archives E081-004 <http://esa.sdsc.edu/Archive/E081-004/e081004d1.html> Accessed 2000 Nov 17.

Rohm C M., Giese JW, Bennett CC. 1987. Evaluation of an aquatic ecoregion classification of streams in Arkansas. Journal of Freshwater Ecology 4(1):127-140.

Scheiner SM. (1993) Introduction: Theories, hypotheses, and statistics, In Scheiner SM, Gurevitch J, (ed.) Design and analysis of ecological experiments. Chapman and Hall, New York.

Šmilauer P, Lepš J. 1999. Multivariate analysis of ecological data. University of South Bohemia Faculty of Biological Sciences, Ceske Budejovice, Czech Republic. <http://regent.bf.jcu.cz/textbook.pdf> Accessed 2000 Sep 27.

Somers KM, Reid RE, David SM. 1998. Rapid biological assessments: How many animals are enough? J North Am Benthol Soc 17:348-358.

Spindler P. 1996. Using ecoregions for explaining macroinvertebrate community distribution among reference sites in Arizona, 1992. Arizona Department of Environmental Quality, Biocriteria Development Program. Phoenix, Arizona, USA.

StatSoft, Inc. 1996. STATISTICA for Windows [Computer program manual]. Tulsa, Oklahoma, USA.

ter Braak CJF, Šmilauer P. 1998. CANOCO reference manual and user's guide to CANOCO for Windows: Software for canonical community ordination. Version 4. Microcomputer Power, Ithaca, New York, USA. 352 pp.

Tetra-Tech 2000. Oklahoma Conservation Commission Data Analysis Plan. Tetra Tech, Inc. Owings Mills MD.

Thorne RSJ, Williams WP, Cao Y. 1999. The influence of data transformations on biological monitoring studies using macroinvertebrates. Wat Res 33(2)343-350.

Taylor, CM, MR Winston, WJ Matthews. 1993. Fish species-environment relationships in a Great Plains river system. Ecography 16:16-23.

Toepfer CS, Williams LR, Martinez AD, Fisher WL. 1998. Fish and habitat heterogeneity in four streams in the Central Oklahoma/Texas Plains Ecoregion. Proc Okla Aca Sci. 78:1998.

U.S. Environmental Protection Agency. 1990. Biological Criteria National Program Guidance for Surface Waters. Environmental Protection Agency, Office of Water. Washington D.C. EPA 440-5-90-004.

U.S. Environmental Protection Agency. 1991. Stream bioassessment technical issue papers: Workshop proceedings. Prepared by EA Mid-Atlantic Regional Operations for U.S. Environmental Protection Agency Office of Wetlands, Oceans, and Watersheds Monitoring Section. Washington, DC.

U.S. Environmental Protection Agency. 1993. An SAB report: Evaluation of draft technical guidance on biological criteria for streams and small rivers. U.S. Environmental Protection Agency, Science Advisory Board Biological Criteria Subcommittee of the Ecological Processes and Effects Committee. Washington D.C. EPA-SAB-EPEC-94003.

U.S. Environmental Protection Agency. 1997. EPA Region 6 Ambient Toxicity Monitoring Program. Dallas, Texas, USA. <http://www.epa.gov/earth1r6/6wq/ecopro/watershd/monitrng/toxnet/index.htm>

Voshell JR, Smith EP, Evans SK, Hudy J. 1997. Effective and scientifically sound bioassessment: Opinions and corroboration from Academe. Human Health Eco Risk Assess 3(6):941-954.

Wang L, Lyons J, Kanehl P, Gatti R. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22(6):6-12.

Wartenburger DS, S Ferson, FJ Rohlf. 1987. Putting things in order: A critique of detrended correspondence analysis. Am Nat 129:434-448.

Whittier TR, Hughes RM, Larsen DP. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Can J Fish Aquat Sci 45:1264-1278.

Yoder CO, Rankin ET. 1995. The role of biological criteria in water quality monitoring, assessment, and regulation. Ohio EPA. Columbus, Ohio, USA. Technical Report MAS/1995-1-3.